The Validity of Judgment:

Can the Assessor Learn to Outperform the Equation?

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

David M. Klieger

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Advisor: Nathan R. Kuncel

September 2010

Acknowledgements

To my advisor, Nathan Kuncel, I owe many thanks. Among them is my appreciation for his sharing with me over the years his vast knowledge of and skill in I-O psychology and psychology in general – its subject matter, research methodology, and practical side. I am extremely grateful for the chance to work with him at the University of Illinois and to continue to work with him at the University of Minnesota and beyond. Thank you for guiding the change in my career path.

I also am indebted to the members of my dissertation committee: Professors John Campbell, Paul Sackett, Mark Davison, and Nathan Kuncel. Thank you for your time and effort while you fulfilled your many other responsibilities. Your insightful feedback made this dissertation a better product.

For their assistance with the dissertation process, I am grateful to other faculty, staff, and several organizations, too. Professor Aaaron Schmidt, Professor Sanford Weisberg, and the Statistical Consulting Service at the University of Minnesota provided extremely helpful guidance on aspects of the research methodology that I employed in my dissertation. Maynard Goff and Personnel Decisions International (PDI) supplied valuable data. The University of Minnesota's Psychology Department awarded me a Graduate Research Partnership Program Fellowship and Summer Research Fellowship to finance several dissertation expenses and to help support my family and me while I labored away at my dissertation. Judy Peterson and Lynn Burchett offered valued guidance as I navigated my way semester after semester around the shoals of teaching and administrative matters while simultaneously working on my thesis.

Also, I would like to express my appreciation to Professors John Campbell, Paul Sackett, Deniz Ones, Joyce Bono, Nathan Kuncel, and Niels Waller at the University of Minnesota as well as Professors Charles Hulin, Fritz Drasgow, Deborah Rupp, Brent Roberts, and Lawrence Hubert at the University of Illinois for helping me to develop much of the knowledge and skill necessary to do the heavy lifting for my dissertation. One could not ask for more accomplished instructors and guides.

From my fellow students at the University of Minnesota, I am lucky to have received their help and friendship. Particular thanks go to Adam Beatty for coding, proofreading and his proficiency with *R*; Matthew Borneman for his statistical advice; Ben Seltzer for his wordsmithing; James Beck for his expertise with longitudinal modeling; Winny Shen for the idea for one of the individual difference variables in my dissertation (gender); Amanda Koch for surrendering her desk to the irresistible force that was my dissertation; and the foregoing graduate students plus Suzie D'Mello, Tom Kiger, Rachael Klein, Mike Natali, Rena Rasch, Jana Rigdon, Kara Simon, and Phil Walmsley for helping me to pilot test my dissertation and for their moral support. Thanks go to also to former graduate students Sara Cooper-Murphy, Brian Connelly, Matthew Borneman, Richard Landers, Sang Woo, and Stephanie Seiler for their friendship along the way. Furthermore, I tremendously appreciate all of the high-quality data entry proofreading, and pilot testing done by undergraduates Qianhui Gong, Derek Burkland, Chelsey Nord, Kayla Leavens, and Andre Hennig.

Last, but certainly not least, I would like to thank my family. To my parents, Iris and Donald Klieger, who sacrificed so much to provide me with opportunities that they never had, and which their own parents could not even imagine. I suspect that I have lived my mother's motto "college is not enough" well beyond expectations. I would like to think that I have made my parents proudest by doing that which is most difficult but most necessary. May they always know how much I love them.

With their frequent visits and care packages, Belle and Arthur Schwartz (my mother-in-law and father-in-law) have provided my family with love and support for which we are forever grateful. I would also like to thank Artie for his incredible handiwork while I have been scribbling away these past few years. Fortunately, he accepts a glass of single malt Scotch as payment. Thanks go to Belle for, among many other things, flying coach with brisket (and not losing it).

I want to express in words my love, admiration, and gratitude for my wife, Karen. I doubt that I can succeed. There are times in my life when I have wondered who would ever put up with me. That she has done so much more than that through so many years of uncertainty and sacrifice is unfathomable. Her love, dedication, and integrity still astound me. She remains, eternally, my "Special K."

As for my son, Maxwell -- what can I say? He is an unbridled joy. I hope that he will, with good reason, think of me as a good father. May he always know how much he is loved and cherished.

Dedication

To Karen and Max.

"There is a prevalent myth that the expert judge of men succeeds by some mystery of divination. Of course, this is nonsense. He succeeds because he makes smaller errors in the facts or in the way he weights them. Sufficient insight and investigation should enable us to secure all the advantages of the impressionistic judgment (except its speed and convenience) without any of its defects."

--Edward L. Thorndike, *Journal of Applied Psychology*, 1918

"About that mystifying enthusiasm…for turning over as many human activities as possible to machinery: What could that have been but yet another acknowledgment by people that their brains were no damn good?"

--Kurt Vonnegut, Jr., *Galápagos*, 1985

ABSTRACT

When judgments (i.e., predictions of outcomes) are incorrect, the negative consequences for individuals, organizations, and society can be serious. For various kinds of outcomes, meta-analyses and literature reviews reveal, time and again, that the predictive validity of information combined in the mind of the assessor ("clinical data combination") is smaller than the predictive validity when the information is combined using an equation or actuarial table ("mechanical data combination"). Therefore, using mechanical approaches instead of clinical ones would seem prudent. However, judgment validity encompasses consequential validity as well as predictive accuracy. Furthermore, even some of the scholars who have emphasized the superior accuracy of mechanical methods admit that it may be possible for a judge to systematically out-predict a mechanical method. One such possible approach is configural reasoning, an assessor's use of a functional form (e.g. an interaction) absent from the mechanical method and yet predictive of the outcome. As indicated by the aforementioned studies indicating the superior accuracy of mechanical combination, judges do not productively employ such techniques in general. Nevertheless, it remains an empirical question whether assessors can be taught to utilize configural reasoning to outperform an equation. In addition, it is important to determine the traits of those individuals who predict and learn to predict most accurately, because identifying such people can minimize the costs of error and training.

This dissertation tries to be comprehensive in scope. It employs experimental designs and methods of assessing individual differences to answer questions about the degree (if any) to which people can be taught to outperform a mechanical equation, the degree (if any) to which assessors can learn to improve the accuracy of their judgments, the degree (if any) to which judges can be made less overconfident in their judgment strategies, the relationship of any changes in accuracy to any changes in confidence, the individual differences that define those

who predict and learn to predict most accurately, and the timing of and extent to which (if any) assessors gain insight about the most accurate predictive approach.

Prior to addressing these issues, this dissertation lays certain groundwork. It clarifies the nomological networks for clinical and mechanical combination. It enumerates much of the vast research that reveals the human cognitive limitations and informational barriers that are thought to contribute to a vicious cycle of lesser clinical accuracy and overconfidence in judgment strategies. Furthermore, it discusses why one should even care about clinical combination if mechanical procedures are generally more accurate.

The most extensive background information provided prior to discussion of the studies conducted by this author concerns the Lens Model as a toolkit for measuring accuracy as well as the determinants of accuracy. Although this portion of the dissertation is somewhat detailed and intricate, it is necessary. First, understanding the Lens Model leads to understanding the determinants of judgment accuracy. Second, understanding the Lens Model leads to understanding how the judge can and cannot outperform the mechanical approach. Third, understanding the Lens Model leads to understanding the limitations of prior research. Fourth, understanding the Lens Model is essential if the reader is to fully understand results, discussion, and conclusions of the author's experiments.

Also reviewed are the "skill score" as an alternative to the Lens Model for measuring accuracy as well as the major considerations involved when teaching people to improve their accuracy and lessen in confidence. The "skill score" provides information about elevation and scatter that is not available from the Lens Model. Final preliminaries focus on experimental design, namely how and why use of a disordinal interaction is central to the experiments conducted by the author, as well as issues concerning the number of experimental cues (predictors) employed, cue redundancy (intercorrelation), the importance of representative design in the experiments, the conduciveness of various types of experimental feedback to learning, and the impact of incentives on judgment accuracy in the experiments.

The author conducted two studies – one in Fall 2009 and another in Spring 2010. Although some of the experimental design details of the studies differed in important ways, their general blueprints were quite similar. Using mostly undergraduate subjects at the University of Minnesota, both studies collected information about individual differences (cognitive ability, gender, personality, interests, and experience). In the experimental portions of the studies, subjects were asked to make predictions of job performance for hypothetical job candidates based on the cognitive ability test score for each candidate as well as how interesting or boring the candidate was expected to find the job. The most accurate clinical prediction strategy would involve applying knowledge that the correlation between cognitive ability and job performance was positive when the applicant was expected to find the job interesting but negative when the applicant was expected to find the job boring (i.e. a disordinal interaction). The competing mechanical model was a linear version of a model that incorporated the disordinal interaction. Subjects were asked about their confidence in how accurately they were making predictions, and in order to assess insight, subjects were asked to narratively self-report the nature of their judgment strategies. Data were analyzed using longitudinal hierarchical linear modeling (for within-person change over time in accuracy, the determinants of accuracy, and confidence), correlation (for between-person differences), and frequencies (mainly for evaluating insight).

Results were fascinating, although many were inconclusive (often due to lack of statistical significance). Although subjects could outperform the mechanical model under certain experimental conditions, this superiority was not statistically significant. Some of the individuals, experimental groups, and/or subject pool means increased or declined in accuracy, the determinants of accuracy, and confidence over time as expected, but often these results were not statistically significant. Nevertheless, there was some evidence that criterion-related feedback about the disordinal interaction led to improved accuracy and decreased confidence while lack of it had the opposite effects. Several individual differences were significantly associated with accuracy, with cognitive ability being the difference most pervasively related to accuracy to a statistically significant degree. Findings for insight were complicated by the inconsistent nature of subjects' narratives. Nevertheless, there was relatively high agreement between raters of subjects' insight, and ratings of insight often had statistically significant correlations with objective measures of accuracy. Moreover, insight as variously measured was often achieved, and if achieved was usually achieved early.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

THE VALIDITY OF JUDGMENT:

CAN THE ASSESSOR LEARN TO OUTPERFORM THE EQUATION?

Overview and Objectives

Poorly-made judgments[1] can have severe consequences for individuals, organizations, and society. Admitting to a school a student who is not prepared for it and who later fails out, hiring an employee who disrupts the workplace, and advancing to CEO a reckless executive who bankrupts the company are just a few examples. Therefore, maximizing the validity of selection systems (i.e., their accuracy for predicting widely-valued outcomes such as job performance and academic performance) is imperative. Extensive literature reviews and meta-analyses consistently demonstrate across different types of outcomes that, as measured by correlation values, the predictive validity of information combined in the mind of the expert (sometimes referred to as "clinical data combination") is smaller than the predictive validity when the information is combined using an equation or actuarial table (sometimes referred to as "mechanical data combination"; Meehl, 1954; Sawyer, 1966; Grove, Zald, Lebow, Snitz, & Nelson, 2000; Karelaia & Hogarth, 2008; Kaufmann & Athanasou, 2009; Kuncel, Klieger, Connelly & Ones, 2010). These findings include the applicant selection context (Kuncel, Klieger, Connelly & Ones, 2010), which is the focal context of this dissertation.

Nevertheless, the very skeptics of clinical data combination have left open the possibility that judges can outperform mechanical data combination. In fact, the promise of clinical judgment over mechanical methods in applicant selection has been discussed at least as far back

[1] Although some people use the terms "judgment" and "decision making" interchangeably, this dissertation will specifically focus on the making of "judgments" rather than "decisions." As Dalal, Bonaccio, Highhouse, Ilgen, Mohammed & Slaughter (2010, pp. 7-8) clarify, "The term *judgment* typically involves a rating or evaluation of the value or attractiveness of a person, an object, an event, a course of action, or some other phenomenon. . . . In contrast, the term *decision* refers to the choice between various alternatives, which may be people, objects, events, courses of action, and so forth." Since the studies described in this dissertation involve making ratings without choosing among the rated targets, "judgment" is the appropriate term. In real-world contexts, a decision (i.e., whom to select) would follow from, and be based upon, the judgments.

as 1918 (see Thorndike, 1918).[2] One suggested means for the judge to systematically outperform a mechanical combination model is through effective engagement in configural reasoning (Meehl, 1959, 1960; Yntema & Torgerson, 1961; Slovic & Lichtenstein, 1971; Einhorn, 1974; Guion, 1998), an approach which will be explained below. Additional possibilities include judges' use of other types of non-linear predictor-criterion relationships (Brannick & Brannick, 1989). Furthermore, the clinician might be able to recognize and employ a brand-new piece of information (in statistical terminology, a new main effect) absent from the mechanical model (i.e., the concept of a "broken leg" discussed in Meehl, 1954). It would appear that, in general, judges do not successfully take advantage of these strategies, since the weight of the review and meta-analytic literature strongly indicates the superior validity of mechanical methods. However, it is an empirical question whether judges can learn to make use of them so as to out-predict the equation.

In the face of contrary empirical evidence, the practice of many judges (e.g., clinicians, assessors, raters, admissions officers, interviewers, hiring managers) reflects a belief that prediction using the clinical combination of data will systematically outperform prediction using the mechanical combination of data (Foderado, 2009; Highhouse, 2008; Jeanneret & Silzer, 1998; Dawes, 1976). In fact, use of clinical data combination to select students is growing (Foderado, 2009). As discussed in greater detail below, suboptimal judgment strategies appear to be both a cause and effect of human overconfidence (Hogarth, 2001; Kuo, 1991; Kahneman & Tversky, 1973; Einhorn & Hogarth, 1978). The extent to which judges can be "debiased" (i.e., made less overconfident), and the relationship (if any) of debiasing to predictive performance, are additional empirical questions (Arkes, 1991).

It is important to avoid overgeneralization of clinicians' predictive accuracy. It is possible that past findings of superior mechanical validities are due at least in part to the failure to

[2] The earliest quantitative analyses of clinical judgment date at least as far back as 1917, when Hughes assessed the accuracy of judges in rating the quality of ears of corn. However, it was Thorndike (1918) who expressed vigorous enthusiasm toward clinical judgment in the selection and assessment context.

sufficiently study individual differences that would identify a subset of people who could systematically predict more accurately than a mechanical method. Even if such a subgroup cannot be identified, it may be possible to train at least some people to regularly outperform the mechanical approach. Given the costs of suboptimal judgments and the costs of training, it is important to identify the characteristics of people who predict, and learn how to predict, best.

Using experimental study designs and various methods of collecting information about individual differences of study participants (cognitive ability and achievement, gender, personality, interests, and experience), I answer six general questions. To what extent (if any) can assessors learn to outperform a mechanical approach in the applicant selection context? To what extent (if any) can assessors learn to improve their judgment validity (in terms of predictive accuracy)? To what extent (if any) can assessors become less overconfident in their judgment strategies? What is the relationship (if any) of any changes in predictive accuracy to any changes in overconfidence? What individual differences (if any) define those individuals who predict, and learn to predict, most accurately? When and to what extent (if any) do individuals possess insight about the optimal strategy for making accurate judgments?

There exists a great deal of empirical literature with the potential to indicate the feasibility and contours of the experimental designs. Therefore, prior to discussion of the experiments and individual differences analyses, there is a discussion of a recent and comprehensive meta-analysis of judgment studies (Karelaia & Hogarth, 2008), an experimental dissertation study very similar to the planned experiment (Camerer, 1981a), and data from real-world applicant selection contexts (summarized in Table 2) to demonstrate the extent to which the planned experiments were both justifiable and feasible. These three studies provide a strong empirical foundation for the current investigation. However, to make sense of this empirical research, it is first necessary to understand specifically what is meant by "mechanical" versus "clinical" data combination, when and why mechanical data combination outperforms clinical data combination, why one should care about the validity of clinical data combination, and the use of the Lens Model and skill scores to measure and explain the predictive accuracy of data combination methods.

LITERATURE REVIEW

Mechanical Data Combination and Clinical Data Combination Defined

In order to clearly understand the mechanical-clinical debate, one must be able to clarify what is meant by "mechanical data combination"[3] versus "clinical data combination"[4]. The distinction between "mechanical data combination" versus "clinical data combination" used here embraces the essence of the definitions in Thorndike (1918), Meehl (1954), and Sawyer (1966), with some additional clarifications. When discussing the difference between mechanical and clinical combination, Thorndike (1918) referred to the use of an equation in comparison to "impressionistic or intuitional use of facts" (p. 75). Meehl (1954) referred to "straightforward application of an equation or table to the data" versus "juggling or weighting . . . done by a clinician" (pp. 15-16). Sawyer (1966) expressly adopted these distinctions and rejected the requirement avowed by Sarbin, Taft, and Bailey (1960) that to be mechanical data combination an approach must be mechanically derived as well as mechanically applied. He wrote that "mechanical combination" includes "any set of rules whose application is objective, whatever mixture of experience and intuition their derivation involves" (p. 180).[5] "Is the clinician

[3] Sometimes "mechanical" in "mechanical data combination" is alternatively referred to as "actuarial", "objective", "quantitative", "formula-based", "inflexible", "hard", "mathematical", "algorithmic", "formulaic", "machine-like", "formal", "statistical", or "scientific". This dissertation will adopt some of these usages from time to time.

[4] Sometimes "clinical" in "clinical data combination" is alternatively referred to as "expert", "human", "impressionistic", "holistic", "comprehensive", "flexible", "soft", "individualized", "in-the-head", "qualitative", "subjective", "informal", "policy-based", "intuitive", "experiential", or "judgmental". This dissertation will adopt some of these usages from time to time. "Clinical" should not denote or imply clinical psychology or any other area of study or practice simply because it contains the word "clinical". It is quite possible that a clinical psychologist (or "clinical" anyone) engages in mechanical data combination and that someone in an area of study or practice that omits the word "clinical" engages in clinical data combination. The term "clinical" is merely an historical usage attributed to Meehl (1954) when he expressed concern about data combination practices of fellow clinical psychologists at the time.

[5] While mechanical data combination does not have to involve mechanical derivation, the mechanical data combinations discussed in this dissertation will be mechanically derived through ordinary least squares (OLS) regression. Unit weighting is an alternative that represents a way to measure importance without resort to mechanical derivation, although unit weighting may require standardization of independent variables so that units of measurement do not influence variable weights (see Ghiselli, Zedeck, & Campbell, 1981). Campbell (1974) observed that no single weighting scheme is best across all situations. Unit weighting can result in better prediction than regression weighting under certain circumstances,

involved?" is the fundamental question (Sawyer, 1966, p. 181). Furthermore, Sawyer distinguished "mode of data combination", which is the focus of this dissertation, from "mode of data collection", which is not. While the key issue for mode of data collection – like for mode of data combination – is the degree of clinician involvement, only the method of data collection refers to the objectivity with which a piece of information used in a judgment is gathered (e.g., a score from a highly objective psychometric test versus a score from a highly subjective unstructured interview). Data combination is a separate issue, and both a mechanically (objectively) collected datum and a clinically (subjectively) collected datum can be either mechanically or clinically combined with other data.

This dissertation further clarifies the distinction between mechanical and clinical data combination. In terms of its form, it is acknowledged that data combination is mechanical only if the combination method can be clearly and specifically defined in mathematical terms or well-defined algorithmic instructions, but now with the additional caveat that such definition must occur before the combination is undertaken with any data. The mechanical-clinical distinction must also be distinguished in terms of process: data combination is mechanical only if it is used consistently[6]. This form and process are synonymous in the sense that any inconsistent

---

because "unit weights (1) are not estimated from the data and therefore do not 'consume' degrees of freedom; (2) are 'estimated' without error (i.e., they have no standard errors); (3) cannot reverse the 'true' relative weights of the variables" (Einhorn & Hogarth, 1975). A smaller initial fit of the regression-based alternative, a larger number of predictors, and a smaller number of predictions (i.e., increasing shrinkage of regression-based validity), larger predictor inter-correlations, and a larger fit of a unit-weight based alternative all are advantageous to the validity of the unit-weighted alternative in comparison to the validity of the regression-based alternative; lower reliability in the criterion measure also advantages the unit-weighted alternative; unavailability of a criterion measure or criterion scores makes use of the regression-based alternative impossible (see Einhorn & Hogarth, 1975). However, it is difficult to generalize the versatile characteristics of mechanical applicant selection systems (see Kuncel, Klieger, Connelly & Ones, 2010), and thus it is imprudent to say authoritatively that applicant selection generally is favorable to just one of the weighting approaches. Furthermore, OLS regression is the typical mechanical approach employed in Lens Model studies (see Cooksey, 1996) which serve as a paradigm for this dissertation. Therefore, mechanical data combination in this dissertation will employ OLS regression.

[6]A consistent approach does permit conditionality. For instance, a consistent approach allows for an interaction, where the relationship between a predictor and an outcome is contingent upon the value of another predictor. However, to be part of a mechanical process, the interaction must be weighed and applied consistently. For instance, if the relationship between Y (job performance) and X_1 (cognitive

application of the method would mean that it no longer was faithful to its clear and specific mathematical definition. It is more difficult to define what constitutes a clinical technique for combining data without reference to mechanical data combination, because by its very nature a clinical approach to data combination is less clear and specific. Even when one models holistic judgment, the model is at best a “paramorphic representation” (Hoffman, 1960). In other words, clinical data combination is a “black box” whose exact nature might be impossible to discern in part because its nature can vary from judge to judge as well as across time and contexts for a single judge (Einhorn, 1986).

As indicated already, one may view mechanical versus clinical data combination as a continuum based on the degree of clinician participation and not simply as a dichotomy. According to Cognitive Continuum Theory, human judgment can be characterized as falling along a cognitive spectrum whose extremes are defined by pure intuition (i.e., clinical) and pure analysis (i.e., mechanical), with “quasi-rationality” in the center (Hammond, 1986; Cooksey 1996). Furthermore, it is possible to mathematically model past clinical data combination and use that model consistently to make future predictions. One can “bootstrap” such a model by regressing clinical judgments on objective cue (predictor) values, and this model is then steadily used in lieu of a clinical process whose form and weights can vary across judges and judgments. A bootstrapped model shares the clinician’s past weighting scheme and the mechanical approach’s clarity and consistency. The bootstrapped model can be characterized as non-clinical, because once solidified, it might not capture the judge’s future predictor weights, predictor selection, or predictor modeling. The validity of the bootstrapped model generally has non-trivially exceeded that of the clinical data combination (Karelaia & Hogarth, 2008; Cooper & Werner, 1990; Dougherty, Ebert, & Callender, 1986; Ashton, 1982; Dawes, 1971; Wiggins & Kohen, 1971; Goldberg, 1970; Kunreuther, 1969; Dudycha & Naylor, 1966; Bowman, 1963), but

ability) is positive if X_2 (how interesting one finds the job) > 3 and negative if $X_2 \leq 3$, then that relationship *always* is positive if $X_2 > 3$ and negative if $X_2 \leq 3$.

the validity of the purely mechanical model (i.e., the model created by regressing the actual criterion values on predictor values) generally has non-trivially exceeded that of the bootstrapped model (Karelaia & Hogarth, 2008; Dawes, 1979; Dawes & Corrigan, 1974; Goldberg, 1970). It seems appropriate to classify a bootstrapped model as falling somewhere between mechanical and clinical data combination.

When and Why Mechanical Data Combination Outperforms Clinical Data Combination

Human Cognitive Limitations

Notwithstanding well-established findings regarding its inferior predictive accuracy, clinical combination of information continues (Foderado, 2009; Highhouse, 2008; Jeanneret & Silzer, 1998; Dawes, 1976). Common objections to the finding of superior actuarial validities and its implications generally have been based on questionable reasoning (as discussed by Highhouse, 2008, and Grove & Meehl, 1996).[7] It has been observed that "reaching truly holistic judgments . . . demands feats of information integration that are more incompatible with our understanding of human cognitive limitations . . . ." (Ruscio, 2003). Barring human hand-calculation error, transcription error, or technological malfunction, mechanical data combination (especially if machine-driven) can operate quickly and with perfect consistency in optimally weighting and mathematically combining large amounts of quantitative information. People may be especially good at selecting, measuring and coding variables (Camerer, 1981a; Dawes, 1979; Sawyer, 1966), guessing near-optimal weights (Dawes & Corrigan, 1974), and instructing, revising, and overriding a machine (Camerer, 1981a; Yntema & Torgerson, 1961). However, even when people consciously try to engage in maximal performance in holistically making predictions[8]:

---

[7] Recently, it has been argued that holistic admissions approaches would "counter[] the wave of grade inflation" (Foderado, 2009). Grade inflation reduces variability, which is a common measure of the extent of inter-individual differences. Holistic judgment is generally less accurate than actuarial combination and typically obscures objective and measurable differences among candidates. Therefore, its adoption would exacerbate the negative impact of grade inflation on the ability to accurately distinguish students based on impartial standards.

[8] Obviously, people may consciously choose to engage in less than maximal performance or to adopt a goal other than maximizing predictive accuracy based on a reported criterion. There is an overlap between (a) the reasons that one could give for consciously choosing to engage in less than maximal performance or

- the weights that they apply are still less optimal than those derived mechanically (Highhouse, 2008; Grove & Meehl, 1996; Dawes, 1979; Dawes & Corrigan, 1974)[9];
- compared to actuarial methods, the application of clinical approaches is less consistent across judgments (Grove & Meehl, 1996; Dawes, 1979), including greater irregularity in weighing, interrelating, and selecting cues (predictors);
- many people do not grasp statistical randomness and variability, which may be a leading cause of inaccurate weighing of information (Hogarth, 1975);
- judges often fail to recognize, understand, or accept the concept of regression to the mean, and after its occurrence they therefore try to modify prediction strategies as if regression to the mean were a controllable phenomenon (Kahneman & Tversky, 1973);
- people make judgments based on a belief in the existence of certain relationships even in the presence of evidence that disconfirms that existence (illusory correlation; Chapman & Chapman, 1971);
- in real-world contexts, people often confuse the redundant (intercorrelated) information that they receive for convergent (independent) information and thus become overconfident in making judgments (Hammond, 2007);

---

to adopt a goal other than maximizing predictive validity based on a reported criterion and (b) possible justifications for accepting lower clinical accuracy when consciously trying to maximize it (e.g., diversity).

[9] Ironically, if people used unit weighting (i.e., simply added everything up) rather than whatever it is that they typically do, then under certain circumstances their weighting schemes could match and even outperform the validity of mechanically-derived, regression-based models (see Dawes & Corrigan, 1974; Einhorn & Hogarth, 1975; Einhorn, 1986). However, people often believe that the weights of predictors are not all equal (sometimes an accurate belief) and that perfect prediction is possible (almost never accurate in real-world situations; Einhorn, 1986; Highhouse, 2008). Empirical evidence shows that even when researchers explicitly tell subjects that perfect prediction is impossible, and that even when researchers generally inform subjects about the probabilistic nature of a judgment task, subjects' judgment behavior does not improve (see Brehmer & Kuylenstierna, 1978, 1980; Johnansson & Brehmer, 1979). People stubbornly view the use of unit weights as an unacceptable introduction of error (Einhorn, 1986). They fail to consider that the introduction of this error might be a reasonable tradeoff against the non-trivial risk that unequal relative weights will reverse when sampling and/or measurement error for data are/is moderate to substantial -- a reversal that obviously cannot occur for unit weights that by definition are all the same (Einhorn, 1986; Einhorn & Hogarth, 1975).

- people frequently depend heavily on a single consideration and then, in making a judgment, fail to adequately adjust in light of new information (adjustment and anchoring bias; Tversky & Kahenman, 1974);
- the specific way in which cue (predictor) information is communicated to people may impact the preferences (i.e., relative weights) that people give to the cues when a purely logical analysis of the information indicates that the relative weights should not vary across different presentation conditions (i.e., framing or reflection effects studied by Kahneman & Tversky, 1981);
- although people can rely on memory aids to expand the amount of information available in their working memories, their working memories generally cannot handle more than seven, plus or minus two, pieces of information for a single mental task (Miller, 1956);
- people engaging in clinical data combination are especially vulnerable to information overload and distraction (Wright, 1974; Hammond, 1978), sometimes leading them to:
    - ignore base rates or other prior probabilities (representativeness heuristic discussed by Tversky & Kahneman, 1974; basing judgments on absolute frequencies rather than relative frequencies as discussed by Einhorn & Hogarth, 1978),
    - ignore data (Wright, 1974),
    - cognitively retrieve only those instances that most easily come to mind (availability heuristic discussed by Tversky & Kahneman, 1974; serial position effects such as primacy and recency discussed by Deese & Kaufman, 1957, and Murdock, 1962), and
    - overly weight negative evidence (Wright, 1974);
- people integrate information more slowly than do machines (Yntema & Torgerson, 1961; Wright, 1974);

- the behavior of people who succeed primarily due to good luck rather than sound judgment sometimes is naively held up as a model on which to base one's own behavior (Kahneman & Klein, 2009);
- people habitually mistake self-fulfilling prophesies for sound judgment strategies (i.e., Pygmalion effect, Hawthorne effect, treatment effects, or wicked environments; Kahneman & Klein, 2009; Hogarth, 2001; Einhorn & Hogarth, 1978; Camerer, 1981a);
- unlike mechanical processes, people engaged in clinical data combination possess a self-image which they may seek to preserve at the cost of predictive precision and accuracy (e.g., confirmation bias; Phillips & Gully, 2008);
- judges might believe (correctly or not) that they will not be held accountable for their mistakes, which might make them less careful (Kahneman & Klein, 2009);
- even if people can at least effectively instruct, revise, and override an algorithmic process (e.g., a machine running regression equations), their vigilance over the algorithmic process may decline over time (the automation bias discussed by Kahneman & Klein, 2009, and Skitka, Mosier & Burdick, 1999, 2000); and
- unlike mechanical processes, people often need corrective feedback to avoid engaging in erroneous, after-the-fact explanations (hindsight biases) that cause them to become overconfident in their judgment-making ability, and such corrective feedback is frequently absent (Fischhoff, 1975; Einhorn & Hogarth, 1978; Hogarth, 1981) or inadequate (Hammond, 1971; Einhorn & Hogarth, 1978; Balzer et al., 1989)[10].

[10] Researchers of judgment who consider themselves a part of the NDM (Naturalistic Decision Making) tradition and those that consider themselves a part of the HB (Heuristics and Biases) tradition generally agree that the opportunity to learn is a necessary precondition for developing skill (Kahneman & Klein, 2009). (Investigators of judgment often fall into one of these two, sometimes discordant camps. NDM investigators typically look to non-quantitative indicators of people who are successful in their respective fields as benchmarks of expertise (e.g., peer consensus), whereas HB researchers generally rely on optimal statistical models instead. Thus, HB investigators tend to employ a more stringent definition of expertise (Kahneman & Klein, 2009)).

As one can see, clinical judgment faces many serious cognitive obstacles. The last of these in the enumerated list, lack of adequate feedback, is particularly troubling. If assessors were to receive adequate feedback, it might be possible for judges to overcome at least some of the other enumerated hurdles. For example, an assessor who realizes the complexity of a task after receiving feedback might write down more information or use mnemonic devices to conserve working memory. Alternatively, a judge might resort to a mechanical process in lieu of a clinical approach. Without knowing that something is wrong, why would an assessor change behavior? Further exploration of the feedback issue follows.

Performance & Overconfidence (Excessive Self-Efficacy)

This lack of objective, didactic feedback together with the subtle and often non-conscious operation of the other influences on judgments can be thought of as an open loop system in which the controller does not receive the necessary information to properly adjust processes to attain the desired end state of better predictive accuracy (see Kuo, 1991, for more information about open loop systems in control theory). The situation is akin to ineffective temperature regulation that results from an automated thermostat failing to receive timely and accurate data about how warm or cool the environment is. Hogarth (2001, p. 89) refers to such a setting where needed feedback is inaccurate, untimely, or deceptive and thereby leads to bad intuitions as a "wicked environment" (the polar opposite of which is called a "kind environment" leading to good intuitions). In the relevant psychological literature for holistic judgment, the cause and effect of this dysfunction is sometimes referred to as the "illusion of validity", a judge's erroneous belief that the use of information of which he or she is aware will lead to a successful prediction (Kahneman & Tversky, 1973; Einhorn & Hogarth, 1978).

Not only can the interplay of informational and cognitive obstacles (e.g., the human susceptibility to information overload plus a lack of feedback) cause clinicians to predict more poorly, but at the same time it can facilitate the development in clinicians of an overconfidence that encourages them in (or that at least fails to discourage them from) remaining subject to the abovementioned, obstructive influences when they make subsequent predictions[11]. The relationship between confidence in one's judgment and accuracy of that judgment is generally low, so a confidence that causes a judge to fail to verify his or her intuition often results in avoidable error (Kahneman & Klein, 2009). For example, fractionated expertise and skill often worsens the already troubled informational and cognitive dynamics of judgments (Kahneman & Klein, 2009). Many professionals possess expertise in only some areas of the larger domains in which they work but often are required to make judgments involving different but related and specialized areas of the larger domain. Due to the superficial similarities between the area of actual expertise and the different area, the professional may wrongly but honestly believe that he or she possesses sufficient expertise in the different area. The resulting overconfidence may exacerbate existing informational obstacles in that the professional fails to even realize that he or she needs to seek out feedback or attend to available feedback in order to make successful judgments.[12] Moreover, even if overconfidence fails to drive maladaptive judgment, a psychological need to justify their original investment in maladaptive behavior may compel people to increase their investment in the behavior even in the face of evidence that indicates that the original behavioral investment was a poor choice (i.e., the phenomenon known as escalation of commitment analyzed by Staw, 1976; a.k.a. the sunk cost fallacy). The psychological need to avoid cognitive dissonance (Festinger, 1957) represents another way to label this need to justify

[11] Overconfidence can be thought of as excessive self-efficacy.

[12] Kahneman and Klein (2009, p. 524) refer to the old adage that a true expert is someone who knows when (and perhaps what) they do not know. In 2003, an organization called the Plain English Campaign gave former United States Secretary of Defense Donald Rumsfeld a Foot in Mouth award for the following comments that he made in a press briefing about "unknown unknowns": "Reports that say that something hasn't happened are always interesting to me, because as we know, there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns — the ones we don't know we don't know" (Steyn, 2003). In *Walden* (1882, originally published in 1854), Henry David Thoreau quotes Confucius as the originator of "To know that we know what we know, and that we do not know what we do not know, that is true knowledge" (p. 20). Apparently, old wisdom is not synonymous with pervasive wisdom.

an original investment in maladaptive behavior and a new phenomenon that may induce further maladaptive behavior. Upon recognizing the sub-optimality of their judgment processes, people may feel an emotional discomfort that causes them to deny the evidence of this sub-optimality and to ignore any future evidence of it.

As discussed, use of mechanical data combination is superior to use of clinical data combination in the sense that over the long run mechanical data combination will lead to more accurate predictive performance, on average, than will clinical data combination (Einhorn, 1986). However, some might look to the paradigm of natural selection to argue that there is a literally fatal inconsistency between human confidence in holistic approaches and the greater accuracy derived from using mechanical methods to predict important outcomes. Human cognitive processes and attitudes are tied to neurobiology and social environments, and neurobiology and social environments influence each other. If human beings underperform an equation when predicting the behaviors of other human beings (among other critical outcomes that can affect reproductive success), then how has the human race physically and socially evolved to its current state?

Several potential explanations exist. It is possible that although the average complexity of successful job and academic performance arguably has increased over the history of *homo sapiens*, the relatively recent Industrial Revolution of the 1700s and current Information Age have accelerated that increase. Successful selection for jobs and education might have been an easier task for hunter-gatherer and agricultural societies, when needed knowledge and skill for species survival were more physical (more concrete) in nature and less cognitively complex (less abstract). Furthermore, as Einhorn and Hogarth (1981) grimly point out, it took 160 million years for dinosaurs to become extinct due to the forces of natural selection, and modern human beings have existed for a mere 2½ million years. On a more optimistic note, evolutionary pressures might result in improved holistic prediction over time.[13] Furthermore, there is a difference between a species' survival and the mean and variability of its quality of life. Suboptimal judgment might not threaten survival of a species, but its consequences can increase suffering and reduce enjoyment. In addition, expertise is not without costs. To the extent of those costs expertise is maladaptive, and lack of expertise is not without its benefits (Arkes, 1991).[14] A "quick and dirty strategy" for judgment may result in more errors than a more careful approach, but the time and effort saved might make it the optimal approach and thus the more adaptive one (Arkes, 1991, p. 487). Also, even if human beings perform at a suboptimal level in making predictions, when the consequences of failure are high human beings can sometimes improve their performance to meet the important challenge (Arkes, 1991). Based on these considerations, the pervasive findings of higher mechanical validities do not necessarily conflict with evolutionary theory.

Why One Should Care About Increasing the Validity of Clinical Data Combination: The Many Sides of Judgment Validity

Although mechanical methods may be derived or initiated by a human being who is subject to many of the same limitations as the clinician, mechanical data combinations are algorithmic (e.g., least squares regression, looking at a specific row and column of a table according to a set of pre-existing instructions), without much (if any) human discretion. Unlike the clinical approach, the mechanical method is consistently applied, so there are far more

[13] The cost-benefit ratio for some human adaptations or maladaptations (such as expertise or lack thereof) may have shifted in recent human history. For example, the fight or flight response might have been adaptive especially for the physically dangerous societies that were more prevalent in the past, but its underlying mechanisms can be maladaptive if they lead to an ulcer after the protracted stress that many people feel in modern society (Archer, 1988).

[14] For instance, the representativeness heuristic hampers expert-level judgment, because it often causes people to erroneously conceptualize probabilities based on how similar a target stimulus is to something else and to ignore base rates (Tversky & Kahneman, 1974). However, use of this heuristic can facilitate quicker judgment that is sufficiently accurate for many contexts (Arkes, 1991).

opportunities for the person combining data clinically to fall prey to the pitfalls enumerated above -- hence, greater actuarial accuracy (criterion-related validity). The utility of a selection or promotion system increases as its criterion-related validity increases, all other things being equal (Taylor & Russell, 1939; Naylor & Shine, 1965; Hunter & Hunter, 1984; Brogden, 1949; Cronbach & Gleser, 1965). In comparison to holistic approaches, mechanical combination methods are generally thought to be less expensive to establish and implement (Foderado, 2009; Grove & Meehl, 1996) as well as more transparent to stakeholders (Highhouse, 2008; Foderado, 2009). One can compellingly argue that if the use of mechanical combination of information is (1) generally more valid for predicting widely used criteria like job and academic performance, (2) more cost effective, and (3) more open, then (all other things being equal) mechanical data combination should be employed to maximize utility.

Does this mean that one should utilize a mechanical approach to combining information? First, it may be possible to improve the accuracy (and thus utility) of clinical prediction so that it exceeds the accuracy (and thus utility) of the actuarial process. This potential is a major justification for conducting the experiments discussed later. Moreover, the increase in criterion-related validity of actuarial judgment may come at a direct cost. Regardless of which data combination method is utilized, whether there is a net gain in utility will depend upon whether the gain in accuracy from using one of the data combination methods over the other exceeds any additional cost of using that method over the other. Although some argue that actuarial data combination is more costly than clinical data combination (e.g., Foderado, 2009; Grove & Meehl, 1996)[15], it is possible that mechanical data combination might be more expensive in some circumstances (Einhorn, 1986).

[15] The University of Washington increased its annual budget by $250,000 when it recently adopted a holistic admissions approach which its director of admissions described as not being the most expensive holistic option that the university could have chosen (Foderado, 2009).

Furthermore, the answer to the question posed above is complicated by the fact that "all other things being equal" rarely, if ever, holds. The validity of judgment is multifaceted and often involves tradeoffs. In addition to considerations of accuracy and economic utility, the validity of using a particular method of data combination – like the validity of using a test – may also reflect risk-related, psychological, legal, ethical, social, political, and practical consequences (Messick, 1980). Even the greater transparency of mechanical methods has both its advantages and disadvantages. Since people vary in their personal values and preferences, their opinions about how best to balance competing interests may vary as well.

The conceptualizations of error and systematicity upon which the superiority of mechanical data combination rests are not the only ones that can be adopted (see Einhorn, 1986). In fact, clinical data combination can result in higher payoffs than mechanical data combination (although usually by increasing some kinds of risk; Einhorn, 1986). For instance, an asymmetric risk might justify the use of a holistic process that errs on the side of false positives or on the side of false negatives. If the potential harm from selecting a candidate who would be a poor performer is much greater than the potential harm from failing to select a candidate who would be a successful performer, then an assessor might look for the unique "Achilles' heel" of each applicant (what Meehl (1954) would call a "broken leg") to reduce overall risk of harm. In high-stakes selection (e.g., hiring of neurosurgeons or explosives technicians), one can imagine such a holistic process actually occurring. Even if it were possible to construct a mechanical procedure that more accurately predicts performance in such contexts[16], the construction of that mechanical procedure might require assessors to select candidates who they distrust in order to determine how "Achilles' heels" are related to actual outcomes. Given assessors' perceptions of the risk of harm, as well as the vocational, legal, and political consequences to the assessor should any candidate's underperformance result in injury to a third party or a third party's property, it seems

[16] Given the seemingly limitless possibilities for "Achilles' heels", an actuarial approach could quickly become overwhelmed by the number of variables or procedures that would need to be incorporated.

unlikely that assessors would permit questionable candidates to perform the job. Without assessments of performance on the criterion or a synthetic validity alternative, creating an actuarial method higher in criterion-related validity than the holistic method could become quite challenging.

Other factors might substantiate the use of impressionistic methods for combining information. A clichéd (but cherished) one is simply the extension of knowledge as its own reward. A more pragmatic reason requires a certain degree of acceptance that at least some clinical data combination should or will persist for the foreseeable future and that improving its predictive validity is a worthwhile goal. Klimoski and Jones (2008, p. 352) opined, "[W]e should not only give weight to technical quality (e.g., reliability, validity) but must also include issues of practicality and political realities as factors." They cite as examples of non-technical, systemic considerations the constraints of HR protocols, workplace accountabilities, and the time value of money. Moreover, there may be "non-predictive benefits of expertise" of clinical data combination such as motivating applicants, reducing anxiety, creating an illusion of fairness, providing a source for blame and liability, and educating others (Camerer, 1981a). Replacing an actuarial selection process with a holistic one might improve applicant reactions and thereby increase the size and diversify the composition of future applicant pools. This kind of benefit is not trivial, because it can be very difficult to successfully dissuade people from the common, naïve, and stubborn belief that perfect prediction is possible in the real world (Einhorn, 1986; Highhouse, 2008; Brehmer & Kuylenstierna, 1978, 1980; Johnansson & Brehmer, 1979), and in mechanical processes the risk of predictive error tends to be more transparent (Foderado, 2009; Highhouse, 2008).[17] In addition, abandonment of clinical data combination might also imperil some assessors' jobs whose existence depends upon its use (Grove & Meehl, 1996).

[17] At least publicly, people tend to espouse honest prediction, too. Consequently, the greater transparency of a mechanical method can create a perverse incentive to use a clinical approach. The greater transparency of a mechanical method makes it more vulnerable to criticism when it is a deliberative process of selecting applicants based on reasons unacceptable to the wider society. For example, the University of

In the context of student selection in the United States, mechanical data combination is at best risky and at worst outright unlawful. In Gratz v. Bollinger (539 U.S. 244, 2003), the ruling of the majority of justices on the United States Supreme Court depended in part upon the fact that a college admissions office gave individualized consideration to only some applicants. This office instead relied on a mechanical combination of ratings to select students, and only on occasion did it give to an applicant individualized consideration that ignored the mechanical prediction. That, combined with the relatively sizeable weight that the admissions office gave to racial minority status, convinced the Court majority to strike down the admissions policy as violating U.S. constitutional and civil rights laws. With regard to the use of mechanical data combination specifically, the majority opined,

> Of course, as Justice Powell made clear in *Bakke,* a university need not "necessarily accor[d]" all diversity factors "the same weight," 438 U. S., at 317, and the "weight attributed to a particular quality may vary from year to year depending upon the 'mix' both of the student body and the applicants for the incoming class," *id.,* at 317–318. But the selection index, by setting up automatic, predetermined point allocations for the soft variables, ensures that the diversity contributions of applicants cannot be individually assessed. This policy stands in sharp contrast to the law school's admissions plan, which enables admissions officers to make nuanced judgments with respect to the contributions each applicant is likely to make to the diversity of the incoming class. See *Grutter* v. *Bollinger, post,* at 337 ("[T]he Law School's race-conscious admissions program adequately ensures that all factors that may contribute to student body diversity are meaningfully considered alongside race in admissions decisions") (279).

In light of the Court's use of and reference to ambiguous language such as "nuanced judgments," "all factors," and "meaningfully considered", it is less than fully clear whether the Court's concern here is over the lack of conditional reasoning (patterning, interaction terms) in the policy's mechanical data combination, the policy's use of the same variable weighting scheme across all applicants, both, and/or perhaps other aspects of the selection policy. In any event, since the Court's holding was based on the policy's differential treatment of applicants based on

Illinois at Urbana-Champaign allegedly employed a formal process of favoring unqualified but politically connected applicants that ultimately led to the resignation of the university's president and the replacement of most of the university's trustees (Foderado, 2009). A more opaque clinical process based on the exact same exact considerations might have continued to go largely unnoticed.

race, it was not merely the mechanical nature of the student selection process to which the Court objected. Therefore, it still might be possible to utilize mechanical data combination in student selection if it does not favor a subgroup of a legally protected demographic classification (particularly race). Given the Court's inclusion of the policy's mechanical nature in its reasoning, the somewhat indistinct nature of the Court's concern with the mechanical procedure, and the chance that any selection policy may differentially impact a protected demographic classification in ways unforeseen, the utilization of mechanical data combination in student selection is not without risk.

In Grutter v. Bollinger, 539 U.S. 306 (2003), the majority of justices on the United States Supreme Court ruled in favor of "flexible assessment" in selecting law school students. Unlike the ruling in Gratz v. Bollinger, this decision does not focus primarily on what an admissions committee *cannot* do. Rather, it concentrates on what an admissions committee *can* do without violating the U.S. Constitution (the Equal Protection Clause), the Civil Rights Act of 1964 (Title VI), or 42 U.S.C. §1981. The Court found satisfactory the clinical data combination practices of the University of Michigan Law School when it wrote that the school

> engages in a highly individualized, holistic review of each applicant's file, giving serious consideration to all the ways an applicant might contribute to a diverse educational environment. *Gratz v. Bollinger, ante,* p. __, distinguished. . . . [I]n the context of individualized consideration of the possible diversity considerations of each applicant, the Law School's race-conscious admissions program does not unduly harm nonminority applicants.

Thus, while Gratz v. Bollinger establishes a legal risk of engaging in mechanical data combination, Grutter v. Bollinger seems to provide a safe harbor for engaging in clinical data combination. It is reasonable to assume that *a priori* knowledge of these Supreme Court opinions will encourage institutions of higher education to engage in clinical data combination and eschew mechanical methods.[18]

[18] In fact, it has been argued that the increasing adoption by universities and colleges of holistic (sometimes called "comprehensive") admissions practices is a preemptive move to avoid violation of the law in a legal climate increasingly hostile to race-based preferences (Foderado, 2009, discussing the views

Legal considerations notwithstanding, the opacity of the clinical process might facilitate greater diversity among a selected pool of candidates. Holistic methods can help decision makers avoid the use of explicit and sometimes controversial quotas while increasing types of diversity that would not otherwise exist (Foderado, 2009). It is a common finding that greater reliance on certain pieces of information (e.g., cognitive ability test scores) in actuarial selection often results in less diversity (e.g., smaller percentages of African-Americans and Hispanics) among those selected even though such reliance may result in the more accurate prediction of commonly used criteria (e.g., work and academic performance; Sackett & Wilk, 1994; Sackett, Borneman, & Connelly, 2008; Ones, Viswesvaran & Dilchert, 2005). In the holistic combination of these pieces of information, the judge is free to privately weigh a piece of information (a) inconsistently across applicants, (b) consistently but less heavily than other pieces of information that would not lessen diversity, or (c) not at all. By employing such approaches, the assessor can minimize or negate the adverse impact (both the social as well as legal concept) that a cognitive predictor would otherwise have on the racial and ethnic diversity of a selected candidate pool.

Perhaps more congenial to empiricists, however, is the consideration that while people may be inferior calculators, they can leverage what they are good at to at least minimize the superiority of mechanical data combination. That is, there might still be "predictive benefits of expertise" (Camerer, 1981a). As already mentioned, people are particularly adept at selecting and coding variables (Camerer, 1981a, Dawes, 1979; Sawyer, 1966), guessing near-optimal weights (Dawes & Corrigan, 1974), and instructing, revising, and overriding a mechanical process (Camerer, 1981a; Yntema & Torgerson, 1961). While people might underperform the mechanical data combination if the predictors and predictor functional forms[19] are identical in

---

of Ward Connerly, a regent of the University of California who was described as an activist against affirmative action).

[19] A predictor functional form is part of the way in which a predictor is related to a particular outcome. A function form is different than a weight (unit, beta, subjective, etc.) assigned to the predictor, because it is the function form that describes or prescribes what geometrically can be thought of as the shape of the

both the mechanical and clinical data combinations, *it is presumptuous to assume that the predictors and predictor functional forms in the mechanical and clinical data combinations will always be identical.*

There might exist circumstances under which such congruency is impractical, impossible, or undesirable. As already discussed, incorporation of "Achilles' heels" ("broken legs") into a mechanical equation could quickly become overwhelming. Also, complex functional forms (e.g., interactions) might explain predictor-criterion relationships well, but their complexity can make them challenging to understand and accept. How often would the typical assessor (e.g., hiring coordinator or admissions committee) be willing and able to update a mechanical equation to include additional information, and how complicated would the typical assessor allow the equation to become? Also, how vigilant would assessors be in modifying or replacing the mechanical process if the KSAOs for successful job or student performance can change, or if the applicant pool can change in regard to performance-relevant characteristics?

predictor-outcome relationship. Typically, this relationship is described algebraically, although the generally opaque nature of clinical data combination might make description of cue functional forms for clinical data combination an exercise in guesswork (another aspect of "paramorphic representation" described by Hoffman, 1960). Also, the nature of the functional form may depend upon the number and nature of any other predictors being used to make the prediction. Possible functional forms may include (but are not limited to) simple linear components only (e.g., X_1), components with exponents other than 0 or 1 (e.g., X_1^2), logarithms (e.g., $\log(X_1)$), interactions (e.g., X_1X_2), determinations of minima and maxima (e.g., $\max(X_1, X_2)$), or any combination of these structures (e.g., $X_1^{X_2}$). For example, predictor X_1 and outcome Y might have a parabolic relationship to each other so that a very low or a very high (but not an average) standing on whatever X_1 measures is associated with the same standing on whatever Y measures. In this case, the functional form of X_1 is X_1^2, because squaring standings on whatever X_1 measures will describe or produce the parabolic relationship between X_1 and Y. If there had been other predictors involved (e.g., X_2, X_3, etc.) then perhaps the functional form for X_1 would not have been parabolic. This latter hypothetical describes a statistical interaction, because the predictors interact such that the effect of X_1 on Y depends on values of one or more of the other predictors (X_2, X_3, etc.).

How We Should Measure the Predictive Performance of Data Combination Methods

Although meta-analyses and literature reviews of the mechanical versus clinical debate have narratively discussed the reasons for the greater predictive validity of mechanical data combination, they generally measure and quantitatively compare only the end products of the combination methods (e.g., Meehl, 1954; Sawyer, 1966; Grove, Zald, Lebow, Snitz, & Nelson, 2000). These end products generally consist of zero-order or multiple correlations (i.e., r or R, respectively) that measure (1) the strength of the clinical data combination as the linear relationship between the clinician's judgments (i.e., the clinician's guesses of the actual criterion values) and the actual criterion values as well as (2) the strength of the mechanical data combination as the linear relationship (usually with weights optimized by least squares regression) between the actual criterion values and the actual values of one or more independent variables. These two end products may be compared to each other for each study sample and set of environmental conditions or averaged meta-analytically across studies to ascertain which combination method produces the higher correlation.

Guion (1991, p. 383) has recommended using a more diagnostic alternative to these customary practices. In fact, the traditional paradigm suffers from at least two limitations. First, if someone wishes to better *understand* the processes and mechanisms that underlie clinical data combination, then it is necessary to measure more than just end products. This observation is especially salient when one desires to better understand specifically how and why mechanical data combination is outperforming clinical data combination. Second, if one wishes to *manipulate* (presumably by experimentation) the features of a prediction task to determine if, how, and why outcomes change (e.g., if, how, and why clinical combination outperforms mechanical combination), then it is extremely helpful to possess metrics of any changes in those mechanisms that underlie clinical data combination. To address these measurement issues, this dissertation employs Egon Brunswik's Lens Model. Given certain limitations of using the r_a

correlation metric of the Lens Model, this dissertation also utilizes an absolute measure of predictive performance known as a "skill score".

The Lens Model (Accuracy as Rank Order Similarity)

The Lens Model, as conceptualized (Brunswik, 1952; Hammond, 1955) and then mathematized (Hursch, Hammond, & Hursch, 1964; Hammond, Hursch, & Todd, 1964; Tucker, 1964; Goldberg, 1970; Einhorn, 1974), provides a very useful framework for understanding why the validity of mechanical data combination surpasses that of clinical data combination and how this outperformance might be diminished, eliminated, or reversed.[20] [21] As Guion (1998, p. 408) observed, "Many research paradigms are used to study judgment and decision making; one, particularly well suited to personnel assessments, is called the *lens model*."[22] Its name derives from the conceptualization of cues (a non-statistical characterization of independent variables) as a lens through which a human judge senses and perceives the criterion in the ecology (environment). As a statistical toolkit, the Lens Model measures the correlation between holistic judgments and actual criterion values. Thus, the Lens Model assesses overall accuracy as rank order similarity. Furthermore, it mathematically decomposes that correlation into that part of

[20] A transparent, mathematical derivation of the standard Lens Model Equation appears in Appendix A of Camerer (1981a).

[21] Castellan (1972), Stewart (1976), and Cooksey and Freebody (1985) have extended the Lens Model for multivariate and other applications, which this dissertation will not pursue. Most lens model studies have employed simpler statistical designs (Cooksey, 1996).

[22] There are at least 14 identified judgment and decision theories in addition to the modern Lens Model paradigm (also called Judgment Analysis). They vary in origin, scope, intended function, judgment competence emphasized, principal concepts, loci of concepts, intended uses of research, conceptualization of uncertainty, loci of uncertainty, model(s) of ignorance implicated, idiographic/nomothetic aspects, diversity in subjects, objects, and tasks, stimulus-object decomposition, and judgment decomposition (Cooksey, 1996; Doherty, 1993; Hammond, McClelland, & Mumpower, 1980; Guastello, 1995, 2002; Guastello, Koopmans, & Pincus, 2009). These theories can be quite incompatible with Judgment Analysis (e.g., Chaos Theory's rejection of normal distribution restrictions and adoption of fractional dimensions of measurement) (Mandelbrot & Hudson, 2006; Guastello, 2002). However, they are not necessarily so. In fact, they can complement Judgment Analysis (e.g., such as when one uses principles of Prospect Theory (a.k.a., Heuristics and Biases Theory) to induce the use of biases and heuristics whose effects on judgment the Lens Model can then be used to measure).

clinical validity attributable to the use of the purely mechanical approach and that part of clinical validity attributable to the use of criterion-related information that is absent from the purely mechanical approach. In addition, Lens Model mathematics provide clear and objective ways of measuring and thus better understanding the challenges of clinical prediction by forcing one to focus on the relationship in the ecology of cues to the criterion (resulting from mechanical data combination), the relationship of the same cues to the clinical judgment (the clinician's estimation of the actual criterion value), the relationship of bootstrapped or policy-captured judgments (based on past clinical judgments) to objective predictions (e.g., based on ordinary least squares regression or unit weighting of cues), and the relationship of the residuals of the bootstrapped model to the residuals of the objective predictions.[23]

Brief history of theoretical development. Although it began with a different focus, the Lens Model has evolved into a theoretical framework for the clinical versus mechanical debate. According to Cooksey (1996), the modern Lens Model paradigm (which Cooksey refers to as "Judgment Analysis") owes its origins not only to Brunswik's ideas of probabilistic functionalism (Brunswik, 1952) but also to Social Judgment Theory (SJT) and Cognitive Continuum Theory (CCT). Brunswik's theories of probabilistic functionalism emphasize the use in research of a naturalistic ecology (i.e., a realistic environment in which a judgment is made) to be treated with the same importance as the organism making a judgment in that ecology (Cooksey, 1996). Brunswik originally developed the Lens Model to study human perception only. Its use for studying human judgment began when Hammond integrated SJT into the model in1955 (Brehmer

[23] The experiments utilize what is sometimes called in Lens Model jargon a "double system design", which looks at both purely mechanical combination of data, purely clinical combination of data, and their relationship to each other. The Lens Model has also been used for single system designs and triple system designs. A single system design is either a purely mechanical combination of data or a purely clinical one (i.e., a policy-capturing study). A triple system design is like a double system design in which there are two different judges. r_a, which represents clinical achievement (predictive validity) in the double system design, represents agreement (correlation) between the clinicians' judgments in the triple system design. Triple design research has taken the form of either (1) interpersonal learning design (IPL) studies which investigate clinicians' ability to learn of or from other clinicians' judgment processes or (2) interpersonal conflict (IPC) studies which examine conflict between judges. For more in-depth discussion, please refer to Cooksey (1996).

and Joyce, 1988, p. 1). SJT research initially employed the Lens Model to further understanding of information integration mechanisms, including understanding of how people identify relevant cues and cue-criterion relationships. SJT research also sought to analyze and improve judgments of the ecology as well as agreement between different sets of judgment policies.[24]

In an attempt to unify human judgment fields, Hammond introduced CCT into the Lens Model framework (Cooksey, 1996). Although it appears that researchers of the clinical versus mechanical debate have not explicitly incorporated CCT into their research, the theoretical relevance of CCT to this debate seems relatively clear, because CCT posits a cognitive continuum whose poles are defined by "pure intuition" and "pure analysis", with "quasi-rationality" in the middle. Hammond later tried to integrate this continuum of cognitive modes with five other dimensions: (1) the feasibility of using these modes to work out social problems, (2) a researcher's or policy maker's degree of control over variables (ranging from experimental intervention to mere representation), (3) the likelihood of interpersonal conflict arising (with analytical modes assumed to produce low conflict risk and intuitive methods high conflict risk), (4) the extent to which cognitive activity is covert (with intuitive methods assumed to be more covert than analytical ones), and (5) the degree to which the theory relevant to the cognitive process emphasizes correspondence (judgment accuracy) versus coherence (logical linkages or internal consistency of judgment) (Cooksey, 1996). In addition, CCT holds that the nature of the task and the passage of time during the task influence the cognitive mode employed by a judge (Hammond, 1986). In other words, it is thought that some types of tasks are more likely to lead to a more analytical judgment approach and others to a more intuitive one. CCT further espouses that judges can selectively employ pattern recognition and functional relations based on judgment task characteristics in order to make their judgments (Hammond, 1988; Cooksey, 1996). However, it remains unclear whether CCT has been used to support the argument in favor of

[24] For more information about Social Judgment Theory and its relation to the Lens Model, the reader is referred to Brehmer and Joyce (1988).

clinical data combination that assessors can use pattern recognition and functional relations to achieve predictive validities that match or exceed the predictive validities of mechanical data combination.

Lens Model components (measuring the determinants of accuracy). The modern and standard Lens Model both pictorially and mathematically describes a series of statistical relationships that permits one to decompose the criterion-related validity of clinical data combination into informative parts. Mathematically, the Lens Model can be thought of in terms of the following equation: $r_a = GR_eR_s + C\sqrt{(1-R_e^2)}\sqrt{(1-R_s^2)}$. The parameters (i.e. variables) to the right of the equals sign in this equation include measurements of the extent to which the holistic judge makes predictions like a purely mechanical model and the extent to which the judge uses information absent from the purely mechanical model (i.e.., new functional forms and "broken legs") to make predictions. Lens Model parameters can be thought of as measures of the determinants of judgment accuracy as measured by r_a. Figure 1 is an illustration of the Lens Model, including indications of which model parameters are represented by different types of feedback that can be given to a clinician about the task environment[25] and/or the clinician's judgment policy. Detailed discussion of feedback will occur later. Definitions and descriptions of the statistical parameters that appear in the Lens Model as well as the statistical relationships that define these parameters appear in Table 1.

In order to clarify what these Lens Model parameters represent and how they are related to each other, they will be applied in the context of the first of the naturalistic clinical versus mechanical examinations of Table 2. In this first examination, there are seven cues being used to predict job fit for workers in a home office as well as first-line and middle management (with no

[25] "Task environment" and "task ecology" refer collectively to the information provided by the following sources: task predictability (R_e, the validity of the mechanical data combination); cue (predictor) weights ($b_{1e} \ldots b_{ke}$); functional forms that describe the relationship between each cue and the criterion; cue ecological validities ($r_{1e} \ldots r_{ke}$); and cue intercorrelations (r_{ij}s). These pieces of information are the Lens Model parameters that can constitute what is sometimes called "task information feedback" (see Appendix A, discussion in the main text that follows, and Balzer et al., 1992).

executives). Scores for each cue were rescaled to fit a standard normal distribution (i.e., z-scored) so that differences among cues in their units of measurement did not affect data combination (see Ghiselli, Campbell, & Zedeck, 1981, for discussion of effective weights).

Mechanical criterion-related validity (R_e). As can be seen in Figure 1 and Table 1, the relationship in the ecology (environment) of cues to the criterion results from objectively relating values of Y_e (criterion values) to values of the cues ($X_1 \ldots X_k$). In the example from Table 2, Y_e = job fit, and $X_1 \ldots X_7$ = the seven cues available to predict job fit (measures of cognitive ability, achievement motivation, responsibility/socialization, leadership/dominance, role play task performance, behavioral interview performance, and assessment center test performance). This mechanical data combination usually takes the form of ordinary least squares regression, which produces weights ($b_{1e} \ldots b_{ke}$). These weights for our example ($b_1 \ldots b_7$), after standardization, are as follows: 0.02, 0.20, 0.11, -0.04, 0.12, -0.08, and 0.11. R_e or R_e^2 quantifies the strength of the relationship between the cues as weighted and the criterion. In other words, it can be thought of as measuring the predictive power of mechanical data combination. After regressing obtained fit values on the seven independent variables, an R_e value of 0.30 is obtained (0.22 after shrinkage to the population-level for 7 independent variables and 161 predictions). Depending on whether one wants to talk about the strength of mechanical data combination for the sample examined or the population from which it comes, the predictive validity for mechanical data combination is either 0.30 or 0.22, respectively.

R_e or R_e^2 also can be thought of as "environmental validity." For a naturalistic environment, mechanical data combination typically explains more variability in outcomes in the environment than does an impressionistic method (see Karelaia & Hogarth, 2008, and its underlying data at http://dx.doi.org/10.1037/0033-2909.134.3.404.supp). For laboratory studies, a mechanical method (often with random error deliberately added) actually is used to simulate the statistical features of the environment. The data underlying the Karelaia and Hogarth meta-analysis (2008; http://dx.doi.org/10.1037/0033-2909.134.3.404.supp) reveal that the higher the validity of the environment (i.e. the more predictable an outcome is based on observed predictor information), the *worse* that clinical data combination does in comparison to the environmental validity (and a mechanical equation that captures that environmental validity). When the gap between the clinical validity and environmental validity is measured with the formula $\sqrt{R_e^{\,2} - r_a^{\,2}}$ (where R_e = mechanical validity and r_a = clinical validity), the correlation between environmental validity and this gap is a substantial 0.92.[26] In other words, as environmental validity increases, so too does the gap in favor of mechanical validity – and steeply. Furthermore, the underlying meta-analytic data indicate that those with some training in the prediction task in question as well as novices almost never out-predict the mechanical equation even when cues are simple (i.e., when the prediction task is linear as opposed to non-linear), environmental validity is high (e.g., 1 > R_e > 0.75, indicating that valid cues exist), and they receive task information feedback, which empirical research shows to be the most effective form of feedback.[27] (A more detailed discussion of feedback appears later in this dissertation.)

Cognitive control (R_s). The relationship of the seven cues ($X_1 \ldots X_7$) to the clinical judgment is estimated by objectively relating values of Y_s (predicted values of the criterion resulting from clinical data combination) to values of the cues ($X_1 \ldots X_7$). This mechanical data combination usually takes the form of ordinary least squares regression, which produces weights ($b_{1s} \ldots b_{ks}$). For each of 13 different clinicians for which sufficient data were available in the

---

[26] Where $R_e > 0.05$, the gap = 0.49; where $R_e > 0.15$, the gap = 0.49; where $R_e > 0.25$, the gap = 0.50; where $R_e > 0.35$, the gap = 0.50; where $R_e > 0.45$, the gap = 0.51; where $R_e > 0.55$, the gap = 0.53; where $R_e > 0.65$, the gap = 0.54; where $R_e > 0.75$, the gap = 0.59; where $R_e > 0.85$, the gap = 0.63; where $R_e > 0.95$, the gap = 0.61. This analysis tried to use all data, but some data were missing from the meta-analytic database. When perfect prediction was possible (i.e., $R_e = 1$), data were omitted, because perfect prediction is essentially impossible in real-world scenarios. As discussed in further detail below, how best to interpret the Karelaia and Hogarth meta-analyses (2008) is debatable.

[27] As measured by $\sqrt{R_e^{\,2} - r_a^{\,2}}$ for all feedback conditions, task information feedback results in an increase in the superiority of mechanical validity from 0.35 to 0.45 for novices and a slight decrease in the superiority of mechanical validity from 0.482 to 0.477 for those with some prior training. Under the boundary conditions, data were not available for experts.

example, the judge's clinical estimations of Y_e actual criterion values (i.e., each judge's set of Y_s values) were regressed on the corresponding $X_1 \ldots X_7$ cue values. As mentioned earlier, this procedure is called bootstrapping or policy-capturing, and it provides a mechanical (consistent) process for making future judgments with a weighting scheme derived from the results of clinical data combination.

R_s and R_s^2 measure the validity of the bootstrapped model for predicting the clinical judgment $(r_{\hat{Y}s,Ys})$. That is, they measure the correlation between the predictions of the consistent bootstrapped model ($\hat{Y}_s$) and purely holistic predictions (Y_s). While sometimes referred to in the literature as measuring "cognitive consistency", R_s are R_s^2 are more properly thought of as measuring "cognitive control" over how a clinician implements the clinician's judgment policy (Cooksey, 1996; Hammond & Summers, 1972). For each judge, an R_s value was derived via the regression procedure described, and an overall average weighted by the number of judgments made by each judge was then calculated. The resulting mean R_s was 0.89 and as a multiple correlation represents a high level of cognitive control. Lack of cognitive control $(1 - R_s^2$ or $\sqrt{1 - R_s^2}$) indexes variation in profile judgments in general (variation around $\hat{Y}_j$, where j = the jth profile).[28] In the example using an R_s value of 0.89, lack of cognitive control ($\sqrt{1 - R_s^2}$) equals 0.45.

Some have argued and tried to demonstrate that cognitive control (R_s) is unrelated to clinical validity or achievement (r_a, Hammond, Hursch, & Todd, 1964; Einhorn, 1974). However, across judges, cognitive control and the validity of the clinical combination are

---

[28] This is to be distinguished from lack of cognitive consistency $(1 - R_c^2)$, which indexes variation in repeated judgments for a single profile (variation around $\overline{Y}_j$) and which can be characterized as lack of test-retest reliability (i.e., lack of temporal stability). Cognitive control never exceeds cognitive consistency (i.e., $R_s = r_{Ys,\hat{Y}s} \leq R_c$) (Cooksey, 1996). For the example, no cognitive consistency was determined, because there were no repeated profiles from which to calculate it.

correlated at 0.32 in the example from Table 2, and meta-analytic evidence indicates that the two parameters are substantially positively related.[29] In any case, a low value of R_s may indicate that the judge is using cues and/or cue functional forms absent from the bootstrapped model of the judge's approach (Einhorn, 1974). As will be discussed in more detail below, the large value of R_s for the example (0.89) might indicate that judges are not using unmodeled cues and functional forms, but the large value for R_s in the example might instead indicate a skewed distribution among judges of the assessment workload (see, e.g., Figure 2).

Judgment (clinical) criterion-related validity (r_a). The relationship of clinical judgment (Y_s) to the criterion (Y_e) is the validity of clinical data combination (r_a). It is the result of correlating past clinical judgments of the criterion (i.e., Y_s values) to respective actual criterion values (i.e., Y_e values). In the example, after calculating for each judge these correlations between predicted fit and actual fit and then weighing each correlation by the number of judgments for each judge, the overall mean correlation was found to be 0.13. In the Lens Model literature, this correlation (r_a) usually is referred to as "achievement", and it indexes the predictive strength (validity) of the clinical data combination. It is the basis for the Lens Model Equation $r_a = GR_eR_s + C\sqrt{(1-R_e^2)}\sqrt{(1-R_s^2)}$[30]. In other words, the 0.13 validity of clinical data

[29] The overall reported mean correlation between R_s and r_a was 0.56 for 237 task ecologies analyzed by Karelaia and Hogarth (2008). Based on data underlying Karelaia & Hogarth (2008) and retrieved at http://dx.doi.org/10.1037/0033-2909.134.3.404.supp, the mean correlation was 0.34 for 184 task ecologies in the lab and 0.39 for 65 task ecologies in field studies.

[30] The original version of the Lens Model Equation is $r_a = \frac{R_e^{\ 2} + R_s^{\ 2} - \sum d}{2} + C\sqrt{(1-R_e^2)}\sqrt{(1-R_s^2)}$ where $\sum d = \sum_i (\beta_{e_i} - \beta_{s_i})(r_{e_i} - r_{s_i})$ (Hammond, Hursch, & Todd, 1964). Thus, the original version involved the calculation of the sum of the products of the difference between standardized ecological regression weights (β_{e_i} s) and standardized utilization/judgmental (s) regression weights (β_{s_i} s) and the difference between ecological validities (r_{e_i} s) and utilization/judgmental coefficients (r_{s_i} s). Tucker (1964) replaced $\frac{R_e^{\ 2} + R_s^{\ 2} - \sum d}{2}$ with GR_eR_s to create what has become the standard version of the Lens Model Equation generally used in this dissertation. The origin of the term "matching index" for G might seem clearer when one observes that in the original

combination (r_a) can be decomposed into $GR_eR_s + C\sqrt{(1-R_e^2)}\sqrt{(1-R_s^2)}$. As one can see, the success of clinical data combination is a function of ecological predictability (i.e., the validity of mechanical data combination, R_e), the clinician's knowledge of the ecology (G and C), and the clinician's control in applying that knowledge (R_s).

GR_eR_s is the "mechanical component" of the clinical prediction, and $C\sqrt{(1-R_e^2)}\sqrt{(1-R_s^2)}$ is the "unmodeled component" (traditionally known by the overly restrictive description "the configural component") of the clinical prediction. (The term "linear" is often used in lieu of the term "mechanical" to describe GR_eR_s. However, GR_eR_s may refer to a non-linear mechanical component of r_a.) "Mechanical component" denotes the use of mechanical data combination (regression) to objectively describe the ecology and the judge's policy model (via bootstrapping). The "unmodeled component" includes cues and cue-criterion relationships which are not captured by the purely mechanical model on the ecological side of the Lens Model or by the bootstrapped model on the judgmental side of the Lens Model. When $C\sqrt{(1-R_e^2)}\sqrt{(1-R_s^2)} \approx 0$ (which many studies arrange or assume), $r_a \approx GR_eR_s$. In other words, the 0.13 achievement (i.e., criterion-related validity) of clinical data combination may have linear and unmodeled components, but some researchers might assume without further investigation that this 0.13 level of achievement is due only to the mechanical component. As one can see in Table 2, this assumption would be incorrect due to a 0.06 value of the unmodeled component.

Mechanical knowledge (G). The extent to which the judge employs the mechanical model in his judgments is indexed via correlation by the term G[31], which is the relationship of

---

version of the equation, $\sum d$ essentially measures how well the ecological and utilization weights and validities match.

bootstrapped or policy-captured judgments ($\hat{Y}_s$, based on past clinical judgments) to objective predictions ($\hat{Y}_e$, based on ordinary least squares regression or unit weighting of cues). For the example, one may generate predicted values from the bootstrapped model using the same data used to create this bootstrapped model as well as predicted values from the purely mechanical model using the same data used to create this purely mechanical model. The correlation between these two sets of predicted values results in a value for *G* for each of the 13 judges. The weighted average (based on number of judgments for each judge) of these correlations is 0.52 (mean *G*). This 0.52 indexes the extent to which the clinician's modeled judgment policy was like the purely mechanical model. Although *G* often is referred to as "linear knowledge" (i.e., knowledge of the task) and may reflect the degree to which the clinician uses a linear mechanical model, its value can instead reflect the degree to which the clinician uses a mechanical model that is non-linear (as the term "non-linear" has been defined below to include interactions, polynomials, etc.). (Nevertheless, the mechanical model in the example happens to be purely linear.) It may be best to think of G as the judge's understanding of how a mechanical regression would operate (hence the label "mechanical knowledge"). Furthermore, *G* is independent of cognitive control (R_s) and actual cue validities ($r_{1e} \ldots r_{ke}$; Castellan, 1992; Hammond & Summers, 1972). It has been demonstrated that values for *G* frequently are highly positively skewed and sensitive to the task environment (Castellan, 1992; Stewart, 1994). Stewart (1994) argued that values for *G* can reflect task difficulty as well as knowledge but indicated that *G* can be a meaningful index within a task ecology. One should interpret *G* cautiously, especially when doing so across tasks.

Unmodeled knowledge (C). Bootstrapping might not fully capture the "approach" that the clinician uses in making judgments, and the purely mechanical model might not fully capture

---

[31] For the same meaning, some authors alternatively use r_m instead of *G* (e.g., Naylor & Schenck, 1968; Goldberg, 1970). Other authors use r_m to represent GR_e, which is the validity of the bootstrapped model (i.e., the correlation between the policy-captured judgments $\hat{Y}_s$ and the criterion values Y_e) (e.g., Camerer, 1981a). Therefore, confusion is best avoided by avoiding the r_m notation altogether.

variability in actual criterion values. A residual (respectively, $Y_s - \hat{Y}_s$ and $Y_e - \hat{Y}_e$) can measure each failure, and one can measure the proportion of variability for which the bootstrapped and purely mechanical models fail to account (respectively, $1 - R_s^{\,2}$ and $1 - R_e^{\,2}$). The relationship of the residuals of the bootstrapped model to the residuals of the objective predictions is symbolized by *C* and is alternatively called "unmodeled knowledge", "unmodeled agreement", and "the residual correlation". Traditionally, this index has been known by the overly restrictive description "configural cue use" (Cooksey, 1996). One may calculate *C* by correlating the residuals of the bootstrapped model ($Y_s - \hat{Y}_s$) with the residuals of the purely mechanical data combination model ($Y_e - \hat{Y}_e$). For the example, just as one could generate predicted scores from the same data used to create the purely mechanical and bootstrapped models, one can generate residual scores from the same data used to create the purely mechanical and bootstrapped models. For purposes of the example from Table 2, the mean correlation between these sets of residual scores after weighting by number of judgments made by each judge (i.e., mean *C*) is 0.16. Thus, there exists variability not captured by a linear model of the judge's policy which also is not already captured by the mechanical data combination.

"[T]he magnitude of *C* specifies the extent to which the clinician effectively uses those special properties which distinguish him from the [purely mechanical] multiple-regression equation" (italics removed; Hammond, Hursch, & Todd, 1964, p. 445). In particular, unmodeled knowledge consists of clinicians' reliance on cues or cue functional forms (including cue interactions) that are not modeled in the mechanical equation as well as chance agreement between random model errors (Cooksey, 1996).[32] In other words, *C* represents variability in

[32] As explained in Table 1, the variability of the residual of the bootstrapped model (i.e., $1 - R_s^2$, lack of cognitive control) consists of (1) variability due to absence of cues in the bootstrapped model, (2) variability due to absence of nonlinear cue function forms in the bootstrapped model, (3) variability due to absence of cue interactions in the bootstrapped model, and (4) variability due to random error in the bootstrapped model (Camerer, 1981a; Einhorn, 1974).

clinical judgment that may be predictive of the actual criterion and which is not already captured by the mechanical validity (i.e., 1 - R_e^2, variability that the purely mechanical model has not already used to predict the actual criterion). If the value of C in the ongoing example were not 0.16 but notably smaller or non-positive, then there would be no new variability (due to unmodeled cues and/or cue functional forms) that the judge could leverage to improve his clinical predictions that was not already explained by the mechanical data combination. However, it still remains an open question whether the value of C is large enough to permit the clinician to close or reverse the difference between the validities of clinical and mechanical data combination (those validities being 0.13 and 0.22, respectively).

Criterion-related validity of unmodeled knowledge (r_z). A large enough value of C is necessary but insufficient for the clinician's success over the purely mechanical model. First, whatever unmodeled variability C captures must be positively valid for predicting the actual criterion. Referred to as "residual validity" and symbolized r_z, this validity mathematically is the following portion of the unmodeled component of the Lens Model Equation: $C\sqrt{(1-R_e^2)}$ (Camerer, 1981a; Einhorn, 1974). For the ongoing example, the residual validity equals 0.16, indicating that the clinicians' use of unmodeled sources of variability (possibly from unmodeled cue and cue functional forms) is positively related to the actual criterion (Y_e). Second, it is the unmodeled component of the Lens Model Equation $C\sqrt{(1-R_e^2)}\sqrt{(1-R_s^2)}$ that fully indexes the extent to which unmodeled information contributes to clinical success (r_a). As already mentioned, the unmodeled component equals 0.06 in the example from Table 2, so 0.06 of the 0.13 value for r_a (i.e., about 46% of the validity of clinical data combination) is due to unmodeled variability (possibly unmodeled cues and cue functional forms). The better the purely mechanical model explains variability in the criterion (i.e., the larger the value of R_e) the less that this unmodeled information matters, because it produces a smaller value of $\sqrt{(1-R_e^2)}$ and thus a smaller value of the unmodeled component. Moreover, the more that clinical judgments are under strong cognitive control (i.e., the larger the value of R_s and thus the smaller the value of $\sqrt{(1-R_s^2)}$), the less unmodeled information is available to the clinician.[33] As already indicated, a small value of R_s may signify that the clinician is validly using cues and/or cue functional forms that are not expressly identified (Einhorn, 1974).

A Lens Model-based explanation of why mechanical criterion-related validity exceeds clinical criterion-related validity. One can decompose the Lens Model Equation to elaborate how the validity of mechanical data combination surpasses that of clinical data combination and how this outperformance might be diminished, eliminated, or reversed. That equation again is $r_a = GR_eR_s + C\sqrt{(1-R_e^2)}\sqrt{(1-R_s^2)}$, where r_a represents the criterion-related validity of the clinical data combination, and R_e represents the criterion-related validity of the mechanical data combination. It is instructive to ask when, mathematically, the validity of clinical data combination can*not* surpass that of the purely mechanical model. Such a scenario occurs only when $C = 0$, $R_e = 1$, and/or $R_s = 1$. (None of those three situations arises in the ongoing example from Table 2. The likelihood of their occurrence is discussed starting in the next paragraph. Assume them to be postulates for now.) In that event, the unmodeled component $C\sqrt{(1-R_e^2)}\sqrt{(1-R_s^2)}$ drops out of the Lens Model Equation, and $r_a = GR_eR_s$. As the square root of a coefficient of multiple determination, cognitive consistency (R_s) must have a non-negative value between 0 and 1. Mathematically, G is heavily skewed toward high positive values (Castellan, 1992; Stewart, 1994), and the empirical evidence does not contradict that determination (see data underlying the Karelaia & Hogarth meta-analysis at

[33] Ibid. Camerer (1981a) estimated that about 10% to 15% of the variability in the residual $1 - R_s^2$ is attributable to cue function forms (presumably non-linear) absent from the bootstrapped model and that another 39% is attributable to unreliability. Thus, based on these figures, up to 51% of the variability in the judgment residual is due to cues absent from the bootstrapped model.

http://dx.doi.org/10.1037/0033-2909.134.3.404.supp[34]). The foregoing findings, combined with the fact that as correlations R_s and G cannot exceed 1, lead to the conclusion that the correlation r_a (the validity of the clinical data combination) will almost never surpass R_e (the validity of mechanical data combination) mathematically when $r_a = GR_eR_s$.

The probability that $R_e = 1$ or that $R_s = 1$ is extraordinarily low (if not zero) in truly naturalistic settings, so these situations will be dispensed with first. If the mechanical data combination for the ecology can explain all of the variability in the criterion, then R_e (the validity of the mechanical data combination) equals 1 and as a correlation r_a (the validity of the clinical data combination) cannot surpass it mathematically.[35] If $R_s = 1$, then the judge makes his clinical judgments with perfect control. Such a scenario is, by definition, not clinical data combination. Even with the aid of a machine (e.g., a hand-held calculator), a person can (and inevitably does) make random errors, inconsistently weight cues, inconsistently omit predictive cues, and

[34] The internet-available data underlying the Karelaia & Hogarth (2008) meta-analysis (http://dx.doi.org/10.1037/0033-2909.134.3.404.supp) reveals that non-positive G values existed for only 3 out of 249 task environments (1.2%) containing G values for the overall meta-analysis, only 13 out of 127 (10.2%) of the non-learning or pre-learning task environments containing G values, and only 1 out of 182 (0.55%) of the learning or post-learning task environments containing G values. If one includes only field studies (where it is less likely than in a lab environment that the value of G is artificially constrained or inflated), then non-positive G values existed for only 3 out of 65 task environments (4.6%) containing G values for the overall meta-analysis, 6 out of 26 (23.08%) non-learning or pre-learning task environments containing G values, and none of the learning or post-learning task environments containing G values. Although the interpretation of values of G aggregated across task environments is uncertain (see Stewart, 1994), it is nevertheless mentioned here that the value of G is generally large and positive across task ecologies, with an unweighted mean value (and standard deviation) of 0.80 (0.25 SD) for the task environments in the overall analysis, 0.68 (0.30 SD) for the task environments in the overall field study-only task environments, 0.60 (0.25 SD) for the non-learning or pre-learning task environments, and 0.87 (0.30 SD) for the learning and post-learning task environments.

[35] While this scenario might seem trivial due to its virtual impossibility for real-world applicant selection (and psychology in general), a non-trivial proportion of judgment studies based on the Lens Model have assumed or arranged R_e to equal 1in lab and even field studies. For lab studies, 1 was the value of R_e in 32 out of 168 (19%) of the non-learning and pre-learning task environments and 39 out of 168 (22.2%) of the learning and post-learning task environments analyzed by Karelaia and Hogarth (2008; see http://dx.doi.org/10.1037/0033-2909.134.3.404.supp). For field studies, 1 was the value of R_e in 5 out of 26 (19.2%) of non-learning/pre-learning and learning/post-learning task environments (see http://dx.doi.org/10.1037/0033-2909.134.3.404.supp).

inconsistently fail to account for predictive cue functional forms.[36] Therefore, it is not surprising that neither R_e nor R_s equaled 1 in the ongoing example or the other two naturalistic task environments described in Table 2.

The probability that there is little or no unmodeled knowledge (i.e. that $C = 0$ or at least $C \approx 0$) in applicant selection (and other) contexts remains subject to debate, because its determination requires assessment of the existence of, and ability of judges to adequately recognize and make good use of, cues and cue functional forms missing from the mechanical data combination. A small value of C might be due to the lack of contribution to the residuals of systematic sources (e.g., when there are no unmodeled cues and cue-criterion functional forms that are predictive of the criterion), or it might be due to lack of inter-residual relatedness of such contribution that does exist (e.g., when the judge is unable to recognize or make use of predictive cues and cue-criterion functional forms that do not appear in the mechanical model) (Cooksey, 1996). The ability to explore C and the unmodeled component of the Lens Model has been limited by the many lab and even "field" studies that assume or arrange for the entire unmodeled component $C\sqrt{\left(1-R_e^2\right)}\sqrt{\left(1-R_s^2\right)}$ to equal 0 and the relative shortage of naturalistic explorations of unmodeled judgment.[37]

As will be discussed in greater depth below, it remains possible that unmodeled knowledge exists and can be learned in naturalistic settings, including applicant selection. Take the example from Table 2. Given the overwhelming evidence that mechanical data combination

---

[36] For studies analyzed by Karelaia and Hogarth (2008) with regard to learning, in none of 168 lab study task environments or 26 field study task environments was the value of R_s (cognitive control) equal to 1 (see http://dx.doi.org/10.1037/0033-2909.134.3.404.supp).

[37] For studies meta-analyzed by Karelaia and Hogarth (2008) with regard to learning in a laboratory setting, in 118 out of 168 (70.2%) of the non-learning and pre-learning task environments and 76 out of 168 (45.2%) of the learning and post-learning task environments, C was assumed or constrained to equal 0 (see http://dx.doi.org/10.1037/0033-2909.134.3.404.supp). For so-called "field" studies meta-analyzed by Karelaia and Hogarth (2008) with regard to learning, in 13 out of 26 (50%) of the non-learning and pre-learning task environments and 7 out of 26 (26.9%) of the learning and post-learning task environments, C was assumed or constrained to equal 0 (see http://dx.doi.org/10.1037/0033-2909.134.3.404.supp).

results in higher predictive validities than does clinical data combination, the importance of understanding *C* is decisive, as "the value of this measure represents the only possible advantage the clinician may have over the traditional multiple regression equation. . . ." (Hammond & Summers, 1965, p. 218). The experimental studies later discussed in this dissertation address the possibility that clinicians can sufficiently recognize and utilize unmodeled cue functional forms.

Limitations of existing Lens Model research. As mentioned, unmodeled knowledge might exist and be acquired in naturalistic settings, particularly in the applicant selection context. Use of such knowledge might allow the clinician to outperform the mechanical model. The existing empirical research on this matter is less than clear, but overall it provides at least a glimmer of hope to the clinician. The vast majority of all relevant studies are experimental in nature. In Kareleia and Hogarth (2008)'s meta-analysis of Lens Model learning studies, only 26 out of 194 (13.4%) of the task environments were field studies, and the task properties of some of these field studies are unrealistic (as discussed in greater detail below). Therefore, the external validity of most of the past research is an unavoidable issue.[38]

The very idea for the experimental portion of the studies described below germinated from the limitations of the existing empirical research, most notably Camerer (1981a)'s doctoral dissertation. Empirical studies about selection and placement that have utilized the Lens Model approach have suffered from a very small number of clinicians (e.g., interviewers), lack of criterion information, weak criterion measures, poor clinical accuracy, and failure to look at the validity of the ecological (mechanical) model (see Ganzach, 2000, p. 6, referring to Dougherty, Ebert, & Callender (1986) and Zedeck, Tziner, & Middlestadt (1983)). In addition to its unusually lucid and comprehensive theoretical explanations absent from much of the Lens Model

[38] Contrary to certain beliefs, the issue of generalizability pervades Judgment and Decision Making research and is not limited to just one faction. Researchers who follow the Naturalistic Decision Making (NDM) approach as well as those who follow the Heuristics and Biases (HB) approach conduct studies in both field and lab settings (Kahneman & Klein, 2009).

experimental literature, Camerer (1981a)'s dissertation is the rare Lens Model learning study that does all of the following: utilizes realistic mechanical/ecological validities (R_e values ranging from 0.29 to 0.53), permits unmodeled knowledge (C) to equal a value other than 0, uses correlated predictors, incorporates a disordinal (non-monotone) interaction functional form, and applies quantitative measures of learning to use unmodeled cues and cue functional forms. Consequently, his work is credible and can give the clinician hope of outperforming an actuarial approach.

One can interpret his results as suggesting that clinicians can in fact learn to make better judgments as measured by greater cognitive control (larger R_s^2 values) when a disordinal interaction is present than when it is not. Unfortunately, the conclusions that one can draw from that work are quite limited in scope due to extremely low sample sizes (with no sample size larger than 6 in any experimental condition); experimental conditions that were not investigated at all due to lack of subjects; provision of feedback to subjects that consisted only of outcomes (actual criterion values); failure to provide several Lens Model parameters for the task ecology involving an interaction, thereby making it impossible to compare the clinical-mechanical differential validity for the interaction versus non-interaction task environments; use of judges at a highly selective educational institution who were formally trained in statistics and judgment; and use of an ecological context that is irrelevant to applicant selection (changes in Canadian wheat prices).

Furthermore, even if one is willing to accept the general principle that laboratory and restricted field studies can sufficiently simulate real-world characteristics so as to be generalizable in their findings, it is uncertain how one should assess the existing experimental evidence on the whole. On a study-by-study basis, many researchers have employed statistical significance testing and line graphs to draw conclusions about differences in the values of Lens Model parameters across conditions or time. Such an approach to sampling error risks rejection of positive findings due merely to small sample sizes and arbitrary graphical scaling choices rather than to lack of a meaningful effect in the population (Schmidt & Hunter, 2004). Moreover, alpha levels on which the significance tests are based are arbitrary or due solely to historical convention (Cowles & Davis, 1982), and the *p*-values of significance tests assume as true null hypotheses that rarely are true (particularly the ubiquitous nil variety of null hypotheses; Kline, 2004). In addition, the fact that Lens Model studies typically involve explicit tinkerings with the details of the ecology (e.g., the value of R_e, amount of inter-correlation among predictors, type of learning feedback provided to a judge, etc.) causes results across task ecologies to be unsurprisingly inconsistent.

That raises the question of whether the findings from Lens Model studies should be aggregated, and if so, how. Schmidt and Hunter (1990, 2004) have demonstrated that the dependence of statistical significance testing on sample size makes counting up the number of statistically significant studies versus the number of statistically insignificant studies an inappropriate technique. Besides, the clinical versus mechanical debate is at least an ordinal one, so it is necessary to know the magnitudes of the lens model statistics and not just whether they reach a certain level of statistical significance. The meta-analytic approach of averaging effect sizes and then trying to detect moderators also is insufficient. As Castellan (1992, p.366) wrote, "[T]he various measures like r_a, R_s, and G cannot be compared directly since the ecology or task environment places limits on one measure which are not placed on the others."

In fact, studies vary considerably in many task features, including the nature of the judged targets; the nature of the criterion; the nature of the predictors; the number of predictors; the extent to which the predictors are redundant, if redundant at all; the relationships of the predictors to the criterion (linear, non-linear, polynomial, configural, etc.); how the predictors and criterion are labeled or described (i.e., abstractly or in a meaningful context); whether the ecological validity (R_e) is determined, assumed, or arranged; the value of the ecological validity (R_e); the number of trials per block; and the number of blocks. Other differences include the nature of feedback (if any feedback); the nature of instructions; the individual characteristics of judges, including level of expertise; the number of judgments per judge; and the nature and extent of any incentives (e.g., money) for successful judgmental performance. It is questionable whether a meta-analytic approach of aggregating effect sizes and employing statistical techniques to detect moderators (e.g., SD_ρ, credibility intervals, Q, I^2, τ^2) (Schmidt & Hunter, 2004; Ader & Mellenbergh, 1999; Higgins & Thompson, 2002; DerSimonian & Laird, 1986) would be able to meaningfully sort out the incredible myriad of inter-study differences.

Nevertheless, Kareleia and Hogarth (2008) have attempted to provide meta-analytic evidence regarding the effect of training on the values of Lens Model parameters for 194 different task environments (in addition to descriptive Lens Model relationships for 249 task environments that include many of these 194). Upon close inspection of the meta-analyzed data for the learning studies[39] (available at http://dx.doi.org/10.1037/0033-2909.134.3.404.supp), certain issues arise. First, although a sufficiently large and positive value for C is vital if clinical judgment is to outperform mechanical data combination, the reported overall meta-analytic mean for C (across 204 meta-analyzed learning as well as non-learning studies) was a mere 0.04. However, when one looks at the data underlying the meta-analyzed learning studies, one sees that the value of C equaled 0 in a very large percentage of task ecologies. Additionally, the value of C was missing in 99 out of 194 (51%) of the pre-learning or non-learning task environments and 43 out of 194 (22.2%) of the post-learning or learning task environments. Furthermore, the meta-analysis admits to sometimes assuming a value of 0 for C in order to calculate other Lens Model parameters. The Lens Model parameter estimates of and involving C therefore reflect task environments that may not sufficiently reflect real world possibilities for clinical data combination.

In addition, and as already has been touched upon, the ecological validity (R_e, validity of the mechanical data combination) for many task environments is similarly unrealistic. In 83 out

[39] In the context of learning studies, this dissertation uses the terms "non-learning" and "learning" to describe conditions in a between-subjects experimental design (in which some subjects were trained and others were not), whereas it uses the terms "pre-learning" and "post-learning" to describe conditions in a within-subjects experimental design (in which all subjects performed without training and then later after training).

of 194 (42.8%) of the pre-learning or non-learning task environments and 124 out of 194 (63.9%) of the post-learning or learning task environments, the R_e value is greater than 0.80. These task environments make it mathematically impossible or at least very difficult for the clinician to outperform the mechanical equation. This is not simply due to the fact that R_e is the mechanical validity benchmark that clinical validity must exceed, but also because the unmodeled component of clinical validity on which the clinician needs to rely for success, $C\sqrt{(1-R_e^2)}\sqrt{(1-R_s^2)}$, will mathematically be 0 or very small for such large values of R_e. Moreover, findings from task environments with such large values of R_e lack psychological realism, as they substantially exceed typical levels for applicant selection. The largest average mechanical population-level effect size in previous meta-analytic research on mechanical versus clinical data combination for selection outcomes is 0.58, where the criterion is academic grade point average (see Kuncel, Klieger, Connelly & Ones, 2010, as well as naturalistic data presented below).[40]

The lack of cue redundancy across task environments represents another problem, because naturalistic psychological environments (particularly in applicant selection) may involve inter-correlated predictors (Naylor & Schenck, 1968). This is the case for predictors for the naturalistic applicant selection data discussed below. In the Karelaia & Hogarth meta-analysis of learning studies, only 1 out of every 5 task environments involved any cues that were at least moderately correlated. Cues were exactly orthogonal in 100 out of 194 (51.5%) of the task environments, and cues were only slightly redundant in another 54 out of 194 (27.8%) of the task environments. The absence of much cue redundancy is troubling for generalizability of previous findings, because in the real world people have difficulty correctly identifying cue redundancy (Hammond, 2007), and empirical evidence indicates that in at least some task environments cue

[40] Although the independent variables underlying that multiple correlation generally were not corrected for artifacts (and thus might have resulted in a lower multiple correlation), the multiple correlation of 0.58 does reflect shrinkage to the population.

redundancy impacts clinical judgment differently than does cue orthogonality (e.g., Naylor & Schenck, 1968).

Aside from the presence and use of broken leg countervailers (cues absent from the mechanical combination) the presence and use of non-linear functional forms represents clinicians' only hope for exceeding the validity of mechanical data combination over the long run. However, 140 out of 194 (72.2%) of the meta-analyzed task environments employed cue functional forms that were solely linear (and this exclusive linearity is usually by design). Furthermore, although non-linear relationships in naturalistic applicant selection contexts might be generally unknown, no amount of inductive reasoning can logically lead to a firm conclusion that they do not exist (see Popper & Miller, 1983, regarding the inherent limitations of inductive reasoning). The inclusion of so many linear-only environments makes the meta-analysis seem incapable of demonstrating any evidence that clinical data combination can match or exceed mechanical data combination.

In general, Lens Model parameter values were absent for many of the task environments in the meta-analysis. 68 out of 194 (35.1%) of the pre-learning or non-learning task environments and 11 out of 194 (5.7%) of the post-learning or learning task environments were missing a value for at least one Lens Model parameter other than *C*. Therefore, the meta-analytic Lens Model parameter estimates based on all 194 task ecologies do not derive from the same task environments.

Due to these issues, one who tries to interpret the meta-analytic findings (including those cited in this dissertation) should do so with caution. In the end, only one meta-analyzed learning study (Koele, 1980) utilized task environments that did not clearly have one or more of the aforementioned problems aside from their being laboratory settings, although it is not fully clear whether the cues in those ecologies were orthogonal or not. After the learning treatment, the superior validity of the mechanical data combination (i.e., $\sqrt{R_e^{2} - r_a^{2}}$) in those four task environments changed from -0.10 to -0.43 (i.e., from clinical superiority to an even greater clinical superiority), 0.22 to -0.10 (i.e., from mechanical superiority to clinical superiority), 0.17 to -0.54 (from mechanical superiority to a clinical superiority), and from 0.46 to -0.18 (from mechanical superiority to a clinical superiority). R_s (cognitive control) declined from 0.79 to 0.71 for the first task environment, remained the same for the third task environment (at 0.77), and increased from 0.34 to 0.44 and 0.32 to 0.58 for the second and fourth task environments. It is possible that the decrease in R_s values indicates that the use of cues and/or cue functional forms improved over trials and that this improvement contributed to improved clinical achievement vis-à-vis the validity of the mechanical combination. As will be made clearer below, the fact that the functional forms in these four task environments of Koele (1980) were non-monotone could have given the clinicians additional opportunity that they needed to outperform the mechanical model, which further justifies the rationality for using a non-monotone functional form in the experiments discussed below.

Analysis of real-world data from a private-sector consulting company provides mixed evidence that clinicians can learn to use unmodeled cues or cue form functions to outperform mechanical data combination. The validities for clinical and mechanical data combination, which appear with the rest of the data in Table 2, indicate that there certainly is a need for better clinical prediction. One will notice that in all three data sets, the validity of mechanical data combination (after shrinkage) non-trivially exceeds that of clinical data combination by 0.09, 0.20, and 0.26, respectively. These findings are not very surprising (see Kuncel, Klieger, Connelly & Ones, 2010). More auspicious for clinical data combination is the mean value of *C* of 0.16 in the first data set, indicating that clinicians do possess unmodeled knowledge absent from the bootstrapped and purely mechanical models. Furthermore, a mean r_z (mean validity for the judgment policy residual) value of 0.16 indicates that the unmodeled part of the clinical judgment is positively valid for predicting the actual criterion value. In addition, the unmodeled component of the Lens

Model Equation is 0.06. Therefore, unmodeled knowledge adds to the validity of clinical data combination.

Other results from the naturalistic data are less promising for clinical combination. *C*, the validity of the judgment policy residual for predicting the actual criterion value, and the unmodeled component of the Lens Model Equation are negative for the second and third data sets in Table 2. The negative values signify that use of unmodeled knowledge actually reduces the validity of clinical data combination (a possibility discussed in Hursch, Hammond, & Hursch, 1964). In all three data sets, the values for R_s all are very large. The sizes of this index indicates that there is not much variability from unmodeled cues and unmodeled cue functional forms that the clinician could further leverage after training, because judges are already exercising strong cognitive control over their judgments. That is, their holistic approaches are almost as consistent as a bootstrapped model of their holistic approaches.

However, for all data sets, the large standard deviations for the number of clinical judgments per clinician (10.1 for a mean of 13, 20.7 for a mean of 11.1, and 20.3 for a mean of 25.5, respectively) indicates that the vast majority of judges made very few judgments and that a very few judges made many more judgments. In Figure 2, one can see this skewed distribution of the number of judgments made as well as the decline in the value of R_s as the number of judgments per clinician increases. To achieve a stable judgment policy for each judge, a ratio of 5 cases (i.e. predictions) for each cue is the absolute recommended minimum when employing multiple regression methods (bootstrapping), and that minimum ratio increases as the number of cues and the size of inter-correlations between cues increases (Cooksey, 1996; Stewart, 1988). Looking at Table 2 and Figure 2, one clearly sees that this minimal condition cannot be met for the vast majority of the judges. Therefore, the high mean R_s values may reflect the highly unequal allocation of work to clinicians in these data sets rather than a realistic appraisal of how much unmodeled variability judges can leverage in a naturalistic applicant selection setting when the workload for each judge is more equitable. A reasonable conclusion from these analyses still is that it might be possible for clinicians in real-world applicant selection settings to outperform a mechanical model and that more research is necessary.

Skill Scores (Accuracy as Mean and Standard Deviation Differences, Too)

In addition to using the Lens Model correlation parameter r_a to quantify predictive performance, one can also capture mean and standard deviation differences between sets of scores. Correlations like r_a are limited to describing differences (or similarities) in shape (e.g., the extent to which the rank ordering of Y_e values match the rank ordering of corresponding Y_s values; Cronbach & Gleser, 1953). Therefore, analyses for the experiments in this dissertation make use of what is called a "skill score" (Stewart, 1990; Cooksey, 1996). The skill score is based on mean square errors, distance measures that capture shape, elevation, and scatter information. Conceptually, the skill score is a ratio comparing the judge's accuracy in holistically predicting actual outcomes to the uncertainty of actual outcomes. The skill score (SS) is computed as follows: $SS = 1 - \frac{MSE_Y}{MSE_R}$ where the mean square error (MSE) terms are:

$$MSE_Y = (1/n)\sum(Y_i - O_i)^2 \text{ and } MSE_R = (1/n)\sum(\overline{O} - O_i)^2 \text{ where}$$

n is the number of predictions; Y_i is the i^{th} prediction; O_i is the observed value for the prediction; and $\overline{O}$ is the mean of the observed values (Stewart, 1990). Converting to Lens Model notation: $Y_i = Y_s$; $O_i = Y_e$; and $\overline{O} = \overline{Y}_e$. The skill score can range in value from negative infinity to 1, where 1 indicates perfect prediction. It is common for skills scores of holistic judgment to be negative in value. Although Murphy (1988), Stewart (1990), and Cooksey (1996) propose further refinements to the skill score so that it can be decomposed into more discrete pieces of information, the refinements involve the addition of a measure of shape (r_{YO}) that excludes information provided by the C Lens Model parameter (Cooksey, 1996). Since this C parameter measures the information that a judge needs in order to outperform a mechanical equation, the simpler version of skill score (above) is employed.

It is reasonable to ask if skill score is relevant for applicant selection. While more information is not always better, it is possible that rank-order accuracy (achieving high r_a) is an insufficient goal for a particular selection context. If the objective were solely to select the best of the applicant pool, then maximizing the value of r_a would be enough. However, some academic majors and jobs require that a minimum threshold of performance be satisfied (e.g. engineering majors needing solid comprehension of foundational coursework as measured by first-year grades in order to have any hope of understanding advanced material; nuclear reactor operators meeting the highest safety standards as measured by ratings of the U.S. Nuclear Regulatory Commission). In these cases, picking the best of the applicant pool is not enough. The ability of an assessor to accurately predict elevation -- the ability to minimize the difference between predicted performance and actual performance – is essential. Therefore, skill score that measures elevation (and scatter) is an important dependent variable for applicant selection.

Considerations for Training People To Improve Their Judgment Accuracy

As already discussed, the two ways in which the clinician can regularly outperform the mechanical model are through the use of uniquely predictive nonlinear relationships absent from the mechanical model or uniquely predictive cues absent from the mechanical model (i.e., "broken leg" countervailers described by Meehl (1954)). A main purpose of the experiments discussed in this dissertation was to determine whether certain characteristics of the judgment task, coupled with feedback, would cause the criterion-related accuracy of the clinical data combination to exceed that of the mechanical data combination. At a minimum, the goal was to at least improve clinical accuracy. Therefore, these characteristics should have facilitated (and not impeded) learning. Past empirical research indicates that the following features of a judgment training would meet this didactic goal: the presence of a disordinal interaction, cue intercorrelation, a small number of cues (2) [41], meaningful cue and criterion labels and context, task information feedback (including provision of relative weights for the cues and cue functional forms), and a financial incentive for successful clinical performance. The manipulation in the experiments were with respect to feedback that is provided to subjects who render judgments. More specifically, the manipulation consisted of whether, when, and how subjects were taught about disordinal interactions and were encouraged to consider a disordinal interaction in their predictions of job performance.

Use of a Disordinal Interaction (When Linear Models Do Not Predict So Well)

In the experimental studies, the relationship between cues and the criterion was simulated based on a mechanical functional form (i.e., mechanical model) consisting of a disordinal

[41] Since Smedslund (1955)'s seminal application of the Lens Model to learning, an experimental study like this one is often called a "multiple cue probability learning study" or "MCPL study" (see Cooksey, 1996). If there had been just one cue, then the study could be called a "single cue probability learning study" or "SCPL study" (see Cooksey, 1996).

interaction plus some random error. Subjects making judgments (predictions of performance) were sometimes given feedback about that disordinal interaction which they might be able to use to predict outcomes more accurately than a linear version of the mechanical model. To understand the experimental studies, it is necessary to understand the reason why they are based on a disordinal relationship between cues and the criterion, why the benchmark for mechanical model accuracy is based on a linear version of that disordinal interaction model, and why some of the subjects making judgments are given feedback about the disordinal interaction model. This, in turn, requires an understanding of the robustness of the linear model in explaining variability, even when underlying data are based on a nonlinear model. Without this understanding, one also might fail to appreciate the challenge of developing functional form information that would be absent from a linear mechanical model and that would substantially increment the predictive validity of that linear model.

Given the myriad types of models that can be used to explain the relationship between cues and a criterion, some definitions of prediction models need to be provided to make further discussion clear. "Linear model" is defined here as an additive and compensatory prediction function that excludes cues with exponents other than 0 or 1 as well as configural functional forms. "Configural" is defined here to include what sometimes is called a "patterned relationship" or "patterning" in which a cue has an effect occur only in conjunction with certain effects of one or more other cues, and statistically a configural relationship is an interaction (Lubin & Osburn, 1957). An additive model adds components (independent variables) which can be weighted. The nature of the weighting system (optimal, unit, subjective, etc.) is irrelevant to whether the model is linear. "Additive" has been interpreted to include models with terms that have exponents other than 0 or 1, but such models generally are considered to be nonlinear. In contrast to additive models, non-additive models combine cue effects through means other than addition. "Non-additive" generally has been interpreted to include models with an interaction term because of an interaction's multiplicative nature but with the implicit understanding that the overall model containing the interaction term might involve addition (e.g., $Y = \beta X_1 + \beta X_2 + \beta X_1 X_2$). A compensatory model permits a high score on one independent variable to compensate for a low score on another independent variable. A non-compensatory model does not and thus cannot be additive. A non-compensatory model is nonlinear. Examples of non-compensatory models include conjunctive models (Einhorn, 1970), disjunctive models (Einhorn, 1970), and scatter models (Brannick & Brannick, 1989; Ganzach, 1993, 1994).[42]

For the purpose of the experiments, the mechanical model is a linear model. This is the most commonly used type of mechanical model in judgment, as it even includes simple addition and averaging that people use on a regular basis. Furthermore, and as will be discussed in greater detail, linear models often predict very well in comparison to alternative judgment strategies (Slovic & Lichtenstein, 1971; Dawes & Corrigan, 1974). Nevertheless, due to the existence of the disordinal interaction in the experiments, the clinician may be able to outperform the mechanical model by employing a nonlinear strategy.[43]

---

[42] Conjunctive judgment (a.k.a., a minimum evaluation function; Dawes, 1964) is a multiple cutoff decision procedure that emphasizes the attribute(s) (i.e., predictor(s)) whose value(s) are/is low in comparison to the value(s) of the other attribute(s), and disjunctive judgment (a.k.a., a maximum evaluation function; Dawes, 1964) is a multiple cutoff decision procedure that emphasizes the attribute(s) whose value(s) are/is high in comparison to the value(s) of the other attribute(s) (Ganzach, 1994). (Assume here that all predictors are standardized so that units of measurement do not affect how the size of one predictor's value compares to the size of another predictor's value.) The reason why a conjunctive approach is a multiple cutoff procedure rather than a linear compensatory procedure is due to the requirement that a person possess a certain minimum ability on all the attributes (Einhorn, 1970). A disjunctive approach assesses a person on the person's best ability only (regardless of any other attributes) and thus requires that a person possess a certain minimum ability on only one of the attributes (Einhorn, 1970). A scatter model takes into account the scatter of the predictor scores (i.e., the variability of the scores across the k predictors $X_1, X_2, \ldots X_k$, where $X_1, X_2, \ldots X_k$ are equally important to prediction of the dependent variable) based on the idea that a pattern of predictor scores contains more information than does a weighted sum of the predictor scores (Brannick & Brannick, 1989). An effect of a scatter on judgment would indicate that the clinician is using either a conjunctive or disjunctive judgment rule (Ganzach, 1994). The rule chosen by the clinician depends upon the clinician's goals (Ganzach, 1993).

[43] It is important to recognize that although "mechanical" and "linear" can and have been used to refer to the same judgment approach, they do not mean precisely the same thing. Whether a model is linear, additive or compensatory is distinct from whether a model is mechanical or holistic in combining data. One can consistently employ a clearly and specifically pre-defined linear, non-linear, additive, non-additive, compensatory, or non-compensatory model. Conversely, one can utilize a poorly defined and/or randomly changing approach to combining data that always remains linear or nonlinear, additive or non-additive, compensatory or non-compensatory. As mentioned, the impressionistic nature of a holistic data combination can make it very difficult or impossible to precisely determine the linear, additive, or

Explanations for the utility of using linear models to predict psychological phenomena are psychological, empirical, and mathematical in nature (Hastie & Dawes, 2001). First, the mathematical aspects are discussed since they are the clearest, followed by the empirical and psychological perspectives. A major reason for the predictive robustness of linear models is the frequency with which true cue-criterion relationships in studies of linear model validity are conditionally monotone (Dawes & Corrigan, 1974). "Monotone" (a.k.a. "monotonic") describes a relationship that is either never non-decreasing or never non-increasing, where an increase in the value of the criterion is never associated with a decrease in the value of a cue, and a decrease in the value of the criterion is never associated with an increase in the value of a cue (see Figure 3 for graphical examples). Monotonic relationships do not require that observed relations be exactly linear but merely that they be linearly estimable. For example, as illustrated in Figure 4, a curved logarithmic function (in black) is monotonic and can be approximated by a straight line (in red). Even when the relationship between cues and criteria is not conditionally monotonic, the relationship might be convertible into a monotonic one.[44]

Monotone relationships either lack interaction effects (such that the strength and directionality of the relationships for the entire sample can be summarized by a single straight line) or involve only ordinal (non-crossover) interactions (such that a single straight line can summarize the strength and directionality of the relationships for the entire sample after the strength – but not positive or negative directionality – of the relationship for at least one subsample is altered or ignored). In other words, when there is no interaction or when an existing

---

compensatory aspects of a judge's approach. One can try to fit models to the judgments and see which model predicts the most variability, or one can infer the model's aspects based on improvements in judgment after the judge is trained to make use of a specific functional form on which the cue and criterion values are derived.

[44] For instance, one can linearize an inverted U-shaped function by turning the raw scores of the cue into absolute mean deviation scores (Dawes & Corrigan, 1974) or a skewed distribution via log transformation of raw variables (Norman & Streiner, 2008, pp. 310-311). However, such transformation might obscure or fundamentally alter a conceptual interpretation of cue-criterion relationships in an undesirable way, so linearization might be inappropriate in some cases.

interaction is ordinal, a single straight line can fairly well-approximate a cue-criterion relationship across levels of another cue.

Three examples from Hastie & Dawes (2001), which appear in Figure 5, illustrate the conditions under which linear modeling is most successful. The criterion variable is the amount of nonalcoholic punch consumed, and the cues are the stressfulness of the environment and whether the subject (a prisoner) is an alcoholic. The first example lacks any interaction effect (i.e., there are just main effects) – hence the parallelism of the graphed lines. Therefore, a single straight line could summarize the relationship between the environment and the amount of nonalcoholic punch consumed after collapsing the alcoholic and non-alcoholic subjects into a single group. In the second example, involving a non-crossover interaction, the differently sloped lines can be made parallel to each other without changing the direction of the relationship between the environment and the criterion. After the lines are made parallel to each other, a single straight line could summarize the relationship between the environment and the amount of nonalcoholic punch consumed after collapsing the alcoholic and non-alcoholic subjects into a single group. Although this is achieved after changing the direction of both graphed lines, the changes might be small enough not to create much error in prediction or much misunderstanding of the nature of the environment-criterion relationship across levels of the other cue. In the last example, involving a non-monotonic (a.k.a. crossover or disordinal) interaction, there is no single straight line that can replace the two cross lines without changing the direction of one or both crossed lines. Such a change not only would result in a non-trivial amount of quantitative error but it also would eradicate the conceptual understanding that the relationship between the environment and the outcome is fundamentally different (in direction) for one level of the other cue (alcoholics) than it is for the other (non-alcoholics). Therefore, a linear model is particularly inappropriate when cue-criterion relationships are non-monotonic.[45]

[45] Note that in connection with correlational research on configural cue utilization, Slovic & Lichtenstein (1971, p. 726) stated that "interactions in ANOVA must be viewed with suspicion because the

Why linear models substitute well for relationships involving non-crossover interactions but not crossover interactions can be further clarified with the illustrations in Figure 6 reproduced from Camerer (1981a). Whereas the former illustrations place the criterion along the ordinate and the second cue within the body of the graph, these illustrations place the second cue along the ordinate and the criterion within the body of the graph. Y_e is the criterion, and X_1 and X_2 are the cues. The first graph illustrates ordinal interactions, and the second illustrates disordinal interactions. If one assumes that the scores on the cues are scaled to be positive, then one's focus is on the upper right quadrants of the illustrations. In this quadrant of the ordinal interaction, the curved functions (where the criterion $Y_e = 4$ and then 20), straight lines (in red) provide reasonable approximations. In this quadrant of the disordinal interaction, the best-fitting straight line at each value of the criterion Y_e switches direction at a point (in orange) along the abscissa as well as at a point (in orange) along the ordinate. (There are two orange points, because each of the 2 cues can be considered the moderator of the relationship between the other cue and the criterion). The framing of Y_e values in blue and green shows how the trend (increase or decrease) in criterion values reverses once the threshold value marked by the orange dot along the X_2 axis is crossed – hence the reversal in slope of the red lines. Assuming a reasonable, bivariate normal scatter of cue values, there is no single straight line that would describe well the inter-cue relationships when the interaction is non-monotonic.

Another reason for the robustness of linear models is that optimal nonlinear functions become more linear with increasing error in measurement of the independent variables (Dawes & Corrigan, 1974). In Figure 7 is a depiction of how an increase in this error transforms a rectangular function into a curve that can be estimated by a single straight line. As Dawes and Corrigan (1974) explain, the "L"-shaped boundary is a conjunctive step function for which the independent variables are perfectly measured. As the amount of measurement error for the

---

technique lacks invariance properties under believable data transformations." However, they did not further elaborate or provide any evidence.

independent variables increases, the function begins to increasingly look like a curve. Although a straight line would not well-approximate the "L"-shaped conjunctive step function when its independent variables are measured without error, it does well-approximate the curve that represents this function when there is measurement error in the independent variables. That measurement error in predictors is the norm in applicant selection (and psychological assessment generally) increases the utility of using linear models to approximate what in reality are non-linear relationships.

From an empirical perspective, there simply are not many widely-acknowledged, non-monotonic interactions in psychology, and there are no widely-acknowledged, non-monotonic interactions in applicant selection. The bases for alleged disordinal interactions typically are "verbal claims and selective post hoc data analysis" (Hastie & Dawes, 2001). In the context of psychodiagnosis, Meehl (1957, 1959) and Meehl & Dahlstrom (1960) hypothesized and tried to demonstrate the existence of configural (a.k.a. patterned) relationships in psychopathology (as measured by MMPI profiles) which clinicians could use in their judgments to improve diagnosis over an existing actuarial model. Mathematically, such configurality is the presence of an interaction (Lubin & Osburn, 1957). However, it is unclear whether the configurality in question was ever a disordinal interaction. Furthermore, its existence in nature comes into question given subsequent findings based on the same data that a linear model generally outperformed clinicians and models of their judgment policies (Goldberg, 1970). Using the same data, Ganzach (1998) *was* able to use a non-linear strategy to outperform Goldberg (1970)'s best linear model. However, this strategy employed scatter rules rather than disordinal interactions, and the extent to which these results would reflect actual clinical judgment in real-world contexts is uncertain.

The psychological perspective on the robustness of the linear model is less than fully clear. As already indicated, the way in which a person actually combines information to arrive at a judgment is not necessarily the same as a mathematical model that accounts for a large share of the variability in judgment – hence Hoffman (1960)'s use of the term "paramorphic representation" term to describe models of human judgment. That being said, it has been argued on one hand that the linear model will perform well in comparison to holistic judgment due to many of the human cognitive limitations enumerated earlier in explaining why clinical data combination underperforms mechanical data combination. In addition, research has revealed the difficulty people have in attending to non-comparable features in the environment such that people must shift attention from one feature to another (Hastie & Dawes, 2001). Whatever the true nature of a relationship, judges might need to cognitively simplify the judgment task when combining information from several cues by anchoring judgment on an individual cue and then adjusting their judgment after assessing the individual's standing on, and the distribution and predictability of, each other cue (Hastie & Dawes, 2001; Yntema & Torgerson, 1961). On the other hand, some research indicates that humans do engage in nonlinear, noncompensatory, or both nonlinear and noncompensatory strategies (Einhorn, 1971; Ogilvie & Schmitt, 1979; Brannick & Brannick, 1989). Then again, some studies have questioned the representativeness of non-linear representations of human judgment processes found in the literature, concluding that linear models are superior (e.g., Goldberg, 1971). The conflicting results may be due at least in part to moderating influences of the environment, individual, or both. The linearity of human cognitive processes might depend on the extent to which the attributes of the individual, the task characteristics at hand, or both impact cognitive control (R_s) of the judge (e.g., compare Ogilvie & Schmitt, 1979, to Goldberg, 1971). Furthermore, linear rather than non-linear explanations for human judgment may be explained in part by the finding that non-linear approaches such as conjunctive judgment may be cognitively easier for humans to employ in making judgments but more difficult to model than linear processes (Elrod, Johnson, & White, 2004). The robustness of the linear model might be obscuring true nonlinearity of psychological processes.

Given the consistent finding in literature reviews and meta-analyses of (1) the predictive superiority of mechanical data combination over impressionistic data combination, (2) the robustness of linear models in general and especially for conditionally monotonic relationships, and (3) the goal of determining whether the judge can learn to outperform a linear (mechanical) model, the experimental studies employ a function in which a disordinal interaction term uniquely explained a substantial percentage of the variability in the criterion.[46] The experiments' non-monotonic interaction involving cognitive ability, intellectual engagement and performance (illustrated in Figure 9) is largely theoretical at this point, as the empirical evidence for this interaction is mixed and limited (e.g., see Fisher, 1991; Kerce, 1985; Willings, 1984; Drory, 1982; Hill, 1975). Like in other multiple cue probability learning studies involving non-monotonic relationships, the cross-over interaction is constructed (e.g., Camerer, 1981a). If the clinician is unable to outperform the linear (mechanical) model in this laboratory setting, then it seems less likely that the clinician would be able to outperform it in a naturalistic environment.

Number of Cues

Given that most applicant selection contexts involve several cues, it can be argued that a design with only 2 cues is unrealistic. However, many multiple cue probability learning studies show that subjects tend to use a small number of the available cues (Brehmer & Brehmer, 1988). Furthermore, in a meta-analytic comparison among conditions involving 2, 3, and more than 3 cues, the largest clinical criterion-related validity (r_a), unmodeled knowledge (C), and mechanical knowledge (G) parameter values occurred when there were only 2 cues (Karelaia & Hogarth, 2008, Table 2). At the same time, R_s values were almost identical for the three compared conditions (Karelaia & Hogarth, 2008, Table 2). This finding of higher judge performance for a smaller number of cues held up after separating the effects of cue inter-correlation, number of cues and R_e in a meta-regression analysis (Karelaia & Hogarth, 2008). Therefore, if the goal is to

[46] For the Fall 2009 study, the 125 judgments that the subjects were asked to make (the last 25 of them being repeated profiles that were later eliminated) use data selectively taken from the data (2,000 profiles) randomly generated to reflect a disordinal interaction. The data selection process was not random to the extent that it tried to advantage the clinician by constraining the impact of sampling error. Specifically, the 4 sets of 25 profiles that were selected for the main 100 judgments at least roughly mirror the functional forms (including that of the interaction itself), predictor-criterion correlations, and inter-correlations for the entire 2,000 originally generated profiles. Such steps were undertaken to reduce the cognitive complexity of the task for the clinicians. The steps used to reduce cognitive complexity for the Fall 2009 study were not used for the Spring 2010 study, although the Spring 2010 study tried to reduce cognitive complexity in other ways.

provide conditions that are most favorable to the judge, then it makes the most sense to employ a 2-cue experimental design.

A 2-cue design also is most practical. If one wishes to control for the number of possible interactions that might tax a judge's cognitive resources, then keeping the number of cues down to 2 means that the only interaction that the judge needs to consider is the disordinal interaction which is being specifically taught. Although it might seem unlikely that a judge would employ untaught polynomials of the cues in making his judgments, limiting the number of cues nevertheless has the possible benefit of limiting the number of possible polynomial combinations that a judge might consider. Last, subjects will have finite patience and cognitive resources, so the number of cases (i.e., judgments) must be kept manageable. As it is, the complexity of the experiments (i.e., the need for subjects to recognize and learn to use the disordinal interaction) unavoidably increases the number of cases needed in the experiments (Stewart, 1988). Having a smaller number of cues helps to restrain the necessary number of cases, because a smaller number of cues requires a smaller number of cases (i.e., judgments) to achieve stable statistical results when designing a Lens Model judgment task (Cooksey, 1996; Stewart, 1988). Based on the foregoing considerations, both the Fall 2009 and Spring 2010 experiments employ only 2 cues.

Cue Redundancy

In both the Fall 2009 and Spring 2010 studies, cues are redundant.[47] It will be assumed that in the real world there is a positive correlation between how intellectually interesting work is and the cognitive ability level of the worker based on the theory that workers higher in cognitive ability are more likely to seek out and be assigned to more interesting work. There are tradeoffs in utilizing cue inter-correlation. Meta-analytic evidence indicates that cue redundancy results in lower clinical criterion-related validity (r_a) and mechanical knowledge (G) values (Karelaia & Hogarth, 2008). This finding of lower judge performance for a lower cue redundancy held up

[47] Since judges usually lack precise knowledge of cue inter-correlation values but may possess at least a general sense of cue relatedness, cue inter-correlation is often called "cue redundancy" in the literature of multiple cue probability learning. This terminology sometimes will be used here.

after separating the effects of cue inter-correlation, number of cues, and R_e in a meta-regression analysis (Karelaia & Hogarth, 2008). However, it is debatable whether this analysis is sufficient given the finding that the existence of redundancy was highly skewed across studies – found in 15% of contexts with 2 cues, 39% of contexts with only 3 cues and 84% of contexts with more than 3 cues (Karelaia & Hogarth, 2008).

Moreover, positive cue redundancy (as in the experiments) attenuates the size of the disordinal interaction and makes linear models fit better than they would if cues were orthogonal to each other (Camerer, 1981a). (Note that negative cue redundancy has the opposite effect.) In Figure 8 is an illustration of this phenomenon in which the positive redundancy is represented by the black oval and the orthogonality is represented by the blue circle. Disordinal interaction functions appear within the body of the graph. The yellow-shaded area represents part of the regions in which the linear model is an especially poor fit. The circle (representing orthogonal cues) covers the more poorly fitting areas better than does the oval (representing positive cue inter-correlation). If it is accepted that a stronger disordinal interaction provides the judge with more opportunity to outperform the linear model, then cue redundancy's attenuation of that strength works in favor of the mechanical model.

Nevertheless, there are important justifications for utilizing inter-correlated cues. One of the reasons that redundant cues (i.e., cognitive ability and how intellectually interesting the work is) are chosen for the experiments is to simulate conditions in naturalistic applicant selection contexts. Most applicant selection cues are positively correlated, because predictors tend to be scaled so that more is better. Although some multiple cue probability learning studies intentionally constrain cues to be unrelated to each other, real-world conditions rarely involve cues that all are orthogonal to each other. Furthermore, there is evidence that cue inter-correlation reduces the cognitive demands of the task (Einhorn, Kleinmutz, & Kleinmutz, 1979). In any event, a non-meta-analytic review of the studies of the impact of cue redundancy suggests that the impact of cue redundancy on the performance of judges varies considerably across task environments (Naylor & Schenck, 1968; Lindell & Stewart, 1974; Bremer, 1974; Armelius & Armelius, 1974; Schmitt & Dudycha, 1975; Lindell, 1976; Camerer, 1981a). Based on the foregoing considerations, both the Fall 2009 and Spring 2010 studies employ positively correlated cues.

Representative Design

Task familiarity and task congruency can affect judgment process quality (Cooksey, 1996). Consistent with the principles of Brunswik's probabilistic functionalism (Brunswik, 1952), the designs of the experiments will be representative of the types of judgment tasks encountered in the real-world. However, this representativeness is incomplete. Most subjects' initial familiarity with the task (selecting applicants) will be limited, because the sample will be by necessity a convenience sample from a university undergraduate population. Findings are thus less generalizable than they would be if the subjects were admissions committee members or hiring managers. Furthermore, if judges lack prior experience making their judgments, aspects of the study design that seem inconsequential can produce unstable results (Stewart, 1988, p. 42; see also Cooksey, 1996, p. 91). Stewart and Ely (1984) demonstrated how, due to their inexperience, beginners can fail to consider the full range of cue weights reasonably possible in a real-world setting. Also, neophytes' general mind-sets rather than any concern of theirs for cue tradeoffs tend to govern those weights (Cooksey, 1996). Notwithstanding the foregoing, one could plausibly argue that in a real-world setting experts would be able to draw on their experience to learn at least as well as novices. Of course, previous experience might interfere with learning, too. In any case, the feedback and instructions in the experiments are designed to be thorough and clear for novices.

Additionally, task congruence will be concrete rather than abstract. As Klayman (1988, p. 131) stated, "It helps to know what you're talking about." The units of measurement for the experimental tasks will be like those subjects encounter and assessors use in the real world (e.g., percentiles, scales from 0 to 6), thereby satisfying Brunswik's call for representative design to enhance generalizability of findings. Also, cue and criterion labels will be meaningful (e.g., "performance") rather than abstract (e.g., "variable Y"). This, too, should augment generalizability to the extent that the labels "tell the story" (Klayman, 1988, p. 132). One study that explicitly tested the impact of abstract versus concrete cue and criterion labels found that the use of meaningful labels resulted in higher task knowledge (in this case just mechanical knowledge, *G*) than did the use of abstract labels (Koele, 1980). For abstract versus concrete study contexts, meta-analytic findings saw equal clinical validity (r_a) and mixed but not very different results for other Lens Model parameters (Karelaia & Hogarth, 2008). On the whole, it seems at least reasonable to employ a meaningful environment in the experiments. Ideally, the experiments would employ a scenario that actually exists in naturalistic settings. However, even if the interesting/boring x cognitive ability disordinal interaction (depicted in Figure 9) does not truly exist in reality, its plausibility in the minds of the subjects may be sufficient to facilitate hypothesis generation, hypothesis testing, and learning -- thereby testing this dissertation's research questions (see Klayman, 1988; Camerer, 1981a).

Feedback

As already mentioned in the context of when and why mechanical data combination outperforms clinical data combination, effective feedback is crucial to high clinical accuracy. As an integral part of training, feedback potentially can "debias" (i.e., prevent the use of faulty heuristics and biases in) clinical judgment, assuming that the lower validity of using clinical approaches is due at least in part to the use of faulty heuristics and biases; however, the training feedback needs to be carefully designed (Arkes, 1991).[48] In the experiments, subjects receive cognitive feedback (CFB) that includes task information (TI) only (e.g., R_e for the linear

[48] For example, it can be supposed that an accountant is unlikely to fall prey to the sunk cost fallacy (Arkes, 1991), and at the same time it has been shown that even one or two courses in general economics do not debias students and protect them against the sunk cost fallacy (Arkes & Blumer, 1985). Also, it appears that those trained in advanced psychology better understand the use of control groups than do those trained in advanced chemistry (Lehman, Lempert, & Nisbett, 1988).

mechanical model, R_e for the mechanical model if the interaction term had been included, relative weights for the main effects and the interaction terms, the inter-correlations for the main effect terms and interaction terms, and correlations between a cue with the criterion for each level of the other cue). They do not receive outcome feedback (OFB), cognitive information (CI), functional validity information (FVI) or feedforward (FF).[49] These major feedback types are indicated in the Lens Model display of Figure 1.

This taxonomy of feedback was developed by Hammond et al. (1980) and expanded by Doherty and Balzer (1988). The conceptualizations of FB and FF which are employed follow the "double system design" of the Lens Model in which there are two sides (ecological and judgmental) and just one judge in the Lens Model. Outcome feedback is not cognitive feedback (or vice-versa), but it instead consists of providing the clinician with the actual criterion value (Y_e) after the clinician has tried to predict that value. Cognitive feedback in general does include task information, cognitive information, and functional validity information. Task information describes mechanical relationships between cues, between cues and the criterion, and about ecological predictability (i.e., information about the task) and thus focuses on the ecological side of the Lens Model. Cognitive information describes relationships between cues and the clinician's judgments (i.e., the clinician's cognitive process) and thus focuses on the judgmental side of the Lens Model. Functional validity information describes relationships between aspects of judgments and mechanical aspects of the ecology and takes the form of lens model parameters like r_a, G, and C. Feedforward is information provided before any judgments are to be made, and feedback (FB) is information provided afterwards (Björkman, 1972). FB for a past set of trials is FF for a future set of trials (Doherty & Balzer, 1988).

There is considerable evidence that task information is the crucial feedback element for improving performance in learning tasks (Cooksey, 1996). TI may include task predictability

[49] This statement should be interpreted to mean that the experimental study also will not provide any combinations of feedback, including what sometimes is referred to as "lens model feedback". "Lens model feedback" is feedback that compares TI to CI (Hoffman et al., 1981).

(R_e), cue-criterion validities ($r_{1e} \ldots r_{ke}$), cue weights ($b_{1e} \ldots b_{ke}$), functional forms that describe the relationship between each cue and the criterion, and cue inter-correlations (r_{ij}s) (Balzer et al., 1992). It has generally predicted lens models parameters (including r_a) better than or almost as well as no FB, OFB alone, other CFB types, and CFB combinations (Karelaia & Hogarth, 2008; Balzer et al., 1992). However, meta-analytic regression weights for TI (as well as for all other CFB types and combinations) for predicting C are negative (Karelaia & Hogarth, 2008), which initially might seem problematic to clinical success over the mechanical model since that success depends upon a sufficiently large positive value for the unmodeled component of the Lens Model (i.e., $C\sqrt{(1-R_e^2)}\sqrt{(1-R_s^2)}$). On the other hand, and as already mentioned, many of the studies meta-analyzed by Karelaia and Hogarth (studies on which the independent variable values used in modeling the meta-analytic regression come) may not be generalizable due to their constraints and omissions. Therefore, the meta-regression results should be interpreted vigilantly. Furthermore, it is easier and less costly to obtain and provide just TI than TI with CI, TI with FVI, or TI with CI and FVI (Kessler & Ashton, 1981). Therefore, if TI is alone sufficient for improving learning, then it seems reasonable to exclude other kinds of CFB.

The experimental studies will not provide outcome feedback (OFB), because research indicates that OFB alone is either unhelpful or detrimental to learning or at least less helpful than other feedback types. According to classical test theory, observed scores for a dependent variable contain a random error component (Lord and Novick, 1968). Clinicians may view a judgment task as deterministic, so they might seek impossibly perfect accuracy in a learning process (Highhouse, 2008; Brehmer, 1980). They are unable to know whether an observed outcome value is greater or smaller than its true value, or by how much. Providing OFB across trials might cause judges to "chase error" much as a dog chases its tail. Hammond (1971, p. 904) called OFB learning "stupid" and "slow and inefficient". When people do learn from OFB, they are unable to clearly describe what they have learned. In complex and uncertain tasks, people tend not to learn at all from OFB (Balzer et al., 1989; Doherty & Balzer, 1988; Lindell, 1976). Hammond (1971, p. 904) further explained,

> Outcome feedback can be of little help to the learner, for two reasons: first, providing the correct answer (outcome) after having made a judgment is virtually useless, since outcomes are related to the cues in a complex, multidetermined, and uncertain way. There is no simple, rule-bound connection to be discovered, for the same answer can be produced by various combinations of cues. Conversely, identical combinations of cues can provide different answers. Second, the irreducible uncertainty in the task requires a long series of trials in order to distinguish between that which is nearly regular and that which is wholly accidental. There is no alternative to a long series of trials, if the learner is limited to outcomes as feedback.

What the judge requires is information about the environment (i.e., task information). "[I]ntuitive judgment frequently lacks awareness of environmental effects. Thus, when judgments or actions are evaluated by comparison with outcomes, environmental factors influencing these outcomes may not even be considered" (Einhorn & Hogarth, 1978, p. 409).

Meta-analytic data support these findings. For studies in which different FB types possibly were mixed within treatment conditions, the meta-regression coefficient of OFB for predicting r_a was smaller than that of TI (Karelaia & Hogarth, 2008). For studies in which it was unclear if different FB types were mixed within each meta-analyzed treatment condition, the meta-regression coefficient for OFB for predicting r_a was smaller than TI's meta-regression coefficient (Karelaia & Hogarth, 2008). It was also smaller than TI's meta-regression coefficient (and negative) for predicting R_s and G (Karelaia & Hogarth, 2008). For studies in which an FB type or combination was clearly exclusive to each meta-analyzed treatment condition, the meta-regression coefficient for OFB for predicting r_a, R_s, and G was almost always zero or negative as well as smaller than the meta-regression coefficients of all of the other FB types and combinations (Karelaia & Hogarth, 2008). While there are oft-cited studies that indicate that OFB actually undermines any benefits of CFB (e.g., Doherty & Balzer, 1988; Hammond et al., 1973), meta-analytic data show that OFB combined with TI outperforms all other FB conditions in predicting r_a, R_s, and G (Karelaia & Hogarth, 2008). Nevertheless, for predicting unmodeled knowledge (*C*) the meta-regression weight for combining OFB with TI is even more negative than the negative meta-regression weight for TI alone (Karelaia & Hogarth, 2008). The success of the clinician in the experimental studies hinges upon a sufficiently large positive value for *C*, so this finding militates against using the TI + OFB combination. Based on the weight of evidence, it would not make sense to provide either this FB combination or OFB alone.

As already mentioned, the Fall 2009 and Spring 2010 studies do not provide judges with cognitive information (CI). CI may include information about cognitive control (R_s), cue-judgment validities ($r_{1s} \ldots r_{ks}$), cue judgment weights ($b_{1s} \ldots b_{ks}$), functional forms that describe the relationship between each cue and the judgment, the level ($\overline{Y}_s$) of a clinician's judgments across profiles, and/or the variability (SD_{Ys}) of a clinician's judgments across profiles (Balzer et al., 1992). Contrary to Guion's suggestion that CI would improve judgment accuracy (Guion, 1991, pp. 384-385), empirical evidence indicates that for improving clinical judgment, feedback related to judgment policies is at the very least not the optimal feedback type, might actually be detrimental, and is prohibitively costly. As already stated, a great deal of research indicates that among CFB types and combinations TI alone is best for improving judgmental performance. Based on meta-analytic data (Karelaia & Hogarth, 2008), CI's meta-regression weight for predicting r_a is smallest and has the only negative value among meta-regression weights for FB types and combinations. Its meta-regression weight for predicting *C* is negative (Karelaia & Hogarth, 2008), which could be problematic since the success of clinical data combination relies on *C*. CI either does not increment TI at all or by much, and in some cases appears to cause slight declines in performance (Karelaia & Hogarth, 2008; Balzer et al., 1992). In addition, CI can require a greater investment of time and effort than TI, especially in real-world training contexts. Unlike in the experiments, real-world training can employ TI feedback without the need to compute and provide CI. As Balzer et al. (1992, p. 38) explains,

> The return of CI and FVI to individual judges requires (a) the development of a judgment task with a sufficient number of profiles on which to base a stable

> policy equation, (b) the investment of each judge's time to evaluate the complete set of judgment profiles, and (c) the experimenter (or consultant) to compute CI and FVI indices and return them to the judges. Although microcomputer software is available to generate judgment profiles and return CFB measures to individual judges (e.g., Hoffman, 1987; Rohrbaugh, 1986), there is still the considerable investment of time by the judges (who might be hard pressed or unwilling to find the time, such as physicians or business executives) to evaluate the set of profiles.

No CI feedback is provided based upon considerations of the probability that giving clinicians CI feedback alone or in addition to TI feedback will at best provide little benefit, as well as the costs of providing CI in real-world settings.

Judges in both studies also will not receive functional validity feedback, which consists of lens model parameters like r_a, G, and C that describe the relationship between the task and the judgment strategy. There is no evidence that FVI alone or in conjunction with other feedback information will improve learning. This is largely due to the scarcity of empirical FVI studies. Balzer, et al. (1989) stated that for their study there was no research that determined the unique contribution of functional validity information to learning. The dearth of studies about FVI might be the reason why Karelaia and Hogarth's (2008) meta-analysis categorized functional validity information under an "other types of feedback" category. In comparison to outcome feedback and the other forms of cognitive feedback, the meta-regression weight for "other type of feedback" was lowest (and negative) for predicting r_a, G and C and lower than the meta-regression weight for task information for predicting R_s. The empirical research that does address FVI directly (Balzer et al., 1992) indicates that clinical success (r_a) when FVI is combined with task and cognitive information is less than when task information alone or task plus cognitive information is provided. In fact, this feedback trio did not improve R_s, G or C beyond the levels for task information alone. It should be noted that such research did not examine feedback conditions involving FVI alone or in combination with only task information or with only cognitive information. Moreover, FVI requires knowledge of criterion scores, which may be difficult to obtain or provide in real-world settings. Just as with CI, obtaining and providing FVI prohibitively costs time and money (Balzer et al., 1992, p. 38). Last, it is unclear how easily judges understand functional validity information or how easily they can be taught to interpret it accurately and quickly. If the empirical research cannot support the utility of using FVI, then these costs seem prohibitive.

The Fall 2009 and Spring 2010 studies exclude feedforward information (FF), too. Feedforward is information provided to the assessor before the assessor makes any judgments (as opposed to afterwards, as with feedback (FB)). The empirical findings for FF versus FB are mixed, and they are difficult to synthesize because of (1) the different experimental conditions (e.g., different task properties) within and across studies that compare FF to FB as well as (2) use of null hypothesis significance testing and arbitrarily scaled line graphs (with absence of, and inability to derive, appropriate effect sizes) to draw conclusions about differences. If these issues are ignored, then there is some evidence in support of using FF. For ecologies with linear relationships only and ecologies with nonlinear relationships, Deane et al. (1972), Hammond and Summers (1965), and Summers and Hammond (1966) found that r_a was greater for CFF than for CFB. No differences in Lens Model parameters between CFF alone and CFB appeared in Nystedt and Magnusson (1973). Adelman (1981) found that when he provided CFF before a first block of trials, r_a was greater and remained consistently greater after increasing across subsequent blocks than when he provided CFB alone. In a study conducted by Galbraith (1984), FB resulted in higher R_s than did FF and thereby reduced the value of the unmodeled component of r_a (i.e., $C\sqrt{(1-R_e^2)}\sqrt{(1-R_s^2)}$).

Nevertheless, there is also support for providing FB. Athough Steinmann (1976) and Galbraith (1984) concluded that CFF limited the amount of improvement to which CFB could lead, they did not find higher r_a or G for CFF than for CFB.[50] Lindell (1976) found that in the context of TI, FB encompasses and outdoes FF in the sense that FB transmits the same

[50] However, these two studies provided both FF and FB in the same treatment condition, so it is difficult to compare their efficacy (Balzer et al., 1989).

information that FF does but allows the clinician to compare his policy (i.e., cue weights and functional forms) to those of the task right when he receives TI. Also, since FF by definition would always provide judges with information prior to their making of predictions, FF does not permit an experimental design to measure accuracy of prediction when no TI is known. It thus would be questionable to conclude that experimental groups receiving different types of TI started off the same and that any subsequent differences in outcomes were due to treatment effects. In addition, actual criterion values in real-world selection contexts (e.g., job performance ratings from bosses, grades from professors) are collected only after judgments (i.e., selections) are made. If TI-based training (which requires knowledge of criterion values) is to be used, then it would seem more sensible to examine interventions with FB rather than with FF. Although evidence exists in favor of using FF, FB is used in the experiments. The empirical evidence is not unanimous; considerations of internal and external validity favor use of FB; and since FB is FF for subsequent judgments, one is always free to reconceptualize an FB study as an FF one by considering outcomes for only those blocks that immediately follow FB. Perhaps most importantly, if FF were used then there would be no baseline measurement of judgment accuracy for the experimental task when no information is received.

Relative Weights. As part of the task feedback, subjects will receive relative weight information.[51] For the relative weights communicated in task feedback, the experiments employ the C_{xj} variant of usefulness weights, which are the average increases across sub-models in the coefficient of determination (R^2) due to adding a cue (Budescu, 1993; Johnson, 2000). C_{xj} should not to be confused with the C Lens Model parameter that measures unmodeled knowledge. C_{xj} is easy to calculate in a mere 2-cue or 3-cue prediction and for each cue appropriately considers the

[51] Given that there are only 2 cues in the experiments, use of (and thus providing) relative weight information seems reasonable. If there were many more cues, then it might be argued that such information overtaxes clinicians and that use of unit weighting should be encouraged instead. As already discussed, unit weights can perform at least as well as optimal weights under certain circumstances, so their use is a sensible fall-back strategy (see Campbell, 1974; Dawes & Corrigan, 1974; Einhorn & Hogarth, 1975; Dawes, 1979).

cue's effect alone, the cue's effect conditional on all other cues, and the cue's effect conditional on all subsets of cues (Johnson, 2000).

Empirical evidence supports the provision of C_{xj} information in feedback. Although there is ample research demonstrating that using unit weights sometimes can result in predictions that are at least as accurate as those resulting from the use of optimal regression weights (Campbell, 1974; Dawes & Corrigan, 1974; Einhorn & Hogarth, 1975; Dawes, 1979), that research focuses largely on beta weights which are not the basis for the relative weights that are provided in the experiments. Moreover, several studies have shown that provision of task feedback, which often includes relative weight information about the task ecology, has resulted in significant improvement in judgments (Karelaia & Hogarth, 2008). Therefore, it seems appropriate to give subjects the relative weights in the task feedback. The many other ways to calculate the relative weight of a cue that is correlated with at least one other cue were rejected for several reasons. These include: zero-order validity coefficients (r), b unstandardized regression weights, standardized beta regression weights (β), the product of β multiplied by the r for its respective cue (βr), t-statistics associated with each cue, regular usefulness weights (increases in the squared multiple correlation due to adding a cue), regressions of the criterion on a set of cues that are related to the original cues but orthogonal to each other (δ^2), and a variant of the latter in which the original cues are regressed on the orthogonal approximation of the original cues instead of the orthogonal approximation of the original cues being regressed on the original cues (ε) (Johnson, 2000; Cooksey, 1996).

All of the regression weight indices suffer from various defects. If the cues do not share the same scale of measurement, then comparisons of non-standardized b regression weights are uninterpretable (Cooksey, 1996). Although one may standardize the betas and also take their absolute values to avoid negative weights, cue inter-correlation still causes exaggeration in the regression weight of the variable with the highest zero-order correlation with the criterion, reduction of the regression weights of the other cues, and sizeable differences in regression weights due to even small differences in samples (Johnson, 2000). To maximize accuracy, judges should consider cue redundancy if it exists. Since zero-order correlations (*r*s) do not consider that redundancy, findings that judges tend to match β_{si} judgment policy weights to ecological validites (r_{ei}s) rather than to ecological β_{ei} weights (Armelius & Armelius, 1974) also makes use of β as the task feedback weight unappealing. In light of the research findings for using β to measure the importance of a cue, one would expect βr, which incorporates it, also to be an undesirable metric. This conclusion is supported by empirical analysis of mechanically derived weights (Johnson, 2000) and relatively inferior linear matching performance (measured by G) by judges who receive βr values as task feedback (Schmitt, Coyle & Saari, 1977).

As a measure of cue importance, C_{xj} is preferable to several non-regression weights, too. The zero-order validity coefficient, *t*-statistics associated with each cue, and regular usefulness weights are not being chosen as relative weights, because they fail to capture the cue's effect alone, the cue's effect conditional on all other cues, and the cue's effect conditional on all subsets of cues (Johnson, 2000). δ^2 is rejected as the relative weight, because ε and C_{xj} outperform it empirically, and ε is more conceptually appropriate than δ^2 (Johnson, 2000). Although ε and C_{xj} perform about equally (Johnson, 2000), utilization of ε involves using matrix operations, which can be time-consuming. Johnson (2000)'s recommended application of ε as the relative weight is inapplicable for the design of the experiments, because his recommendation assumes that there are at least several cues. If that were the case, then the C_{xj} variant's requirement to average across 2^p-1 sub-models (where p = the number of cues) would be cumbersome. However, given the use of only 2 cues in the experiments, this is not an issue.

Graphical Task Feedback. Feedback in the experimentys was graphical in nature given that this mode is thought to result in better performance than alternatives (Cooksey, 1996; Hammond, 1971). It also appeared in traditional paper-and-pencil format rather than in some automated one (e.g., computerization or slide show). There is some evidence that providing information tonally and permitting the judge to control the cue characteristics (allowing for continuous feedback) results in even higher clinical achievement (Hoffman et al., 1981). Given the rarity of that experimental design in the published literature, its risks and complexity are too uncertain. Furthermore, a traditional paper-and-pencil format in group sessions offers an effective and economical way to monitor a large number of undergraduate participants for good faith performance in the experiments.

Incentives

The nature and amount of any effect of incentives on the validity of clinical data combination in this dissertation's experiment was at best uncertain. Assuming that lower clinical criterion-related validity stems at least in part from clinicians' self-assured reliance on faulty heuristics, biases, and limited cognitive capabilities, it could be reasonably supposed that increases in the criterion-related validity of holistic judgment would result from the "debiasing" of clinical judgment (Arkes, 1991). Most motivation theories imply that the use of incentives would improve clinical accuracy to the extent they reduce biases in favor of suboptimal prediction strategies. In particular, expectancy theory (Lawler, 1971, 1973; Vroom, 1964), reinforcement theory (Komaki, Coombs, & Schepman, 1996), goal-setting theory (Locke, Latham, & Erez, 1988), and equity theory (Adams, 1963, 1965) would suggest that an incentive would increase the performance of judges combining data clinically. Motivation is a major determinant of job performance (Campbell, 1990), including the job performance of those in naturalistic work settings who combine data to select applicants. At least some of those individuals presumably receive performance-based compensation in the form of a commission, bonus, or raise as motivation, so it would seem reasonable to financially incentivize subjects based on their performance, too. At the same time, the extent to which subjects would be intrinsically motivated to exert more than minimal effort effort is unknown. On the other hand, cognitive evaluation theory would suggest that an extrinsic reward would work against any intrinsic motivation (Deci, 1971, 1972a, 1972b; Deci & Ryan, 1985). In addition, meta-analytic evidence suggests that financial incentives do not improve performance quality much in experimental simulations ($\hat{\rho}$ = 0.08, k = 6, N = 351; Jenkins, Mitra, Gupta, & Shaw, 1998). Reported sampling error is small (0.017).

Some researchers have adopted a multi-perspective approach to understanding whether incentives will improve judgment. As Einhorn and Hogarth (1981, pp. 57-58) explain,

> First, it has been claimed that such effects [of persistent judgment errors and biases] can be overcome by increasing incentives (through higher payoffs and/or punishments). In one sense, this argument is irrefutable since it can always be claimed that the incentive wasn't high enough. However, direct evidence shows that increased payoffs do not necessarily decrease extreme overconfidence (Fischhoff, Slovic & Lichtenstein, 1977) nor prevent preference reversals (Grether & Plott, 1979). Furthermore, the indirect evidence from clinical judgment studies in naturally occurring settings, where payoffs are presumably high enough to be motivating, continues to indicate low validity and inferiority to statistical models (Dawes, 1979). In addition, claims that people will seek aids and/or experts when the stakes are high (Edwards, 1975) are predicated on the assumptions that: (a) people know that they don't know; and (b) they know (or believe) that others do. On the other hand, it is foolish to deny that payoffs, and thus motivation, have no effect on processes of judgment and choice. Indeed, one only needs to recall the fundamental insight of signal detection theory (Green & Swets, 1966), which is that both cognitive and motivational components affect judgment (also see Killeen, 1978).

In the experiments, the extent to which subjects know what they do not know and believe that feedback will fill vital gaps in their understanding is unclear. Arkes (1991) posits that whether an incentive will improve judgment accuracy depends upon the nature of the clinical judgment error that would occur without the incentive. An incentive may reduce strategy-based errors that result from judges' failures to adequately pay attention to data (i.e., judges' laziness). On the other hand, incentives might not only fail to reduce association-based errors, but they instead might encourage overconfident judges who make such errors to use their suboptimal strategies with greater vigor (Arkes, 1991). (This is similar to Einhorn and Hogarth (1981)'s concern that a judge needs to know what the judge does not know in order to avoid making errors.) According to Arkes (1991), incentives do not reduce psychophysically-based errors either[52], because incentives do not change the location of the judge's position on the judge's response function or

---

[52] Psychophysically-based errors result from behaviors based on a nonlinear (e.g., curved) relationship between external stimuli and psychological responses. As a result of these behaviors, judges do not accurately respond to extreme stimulus magnitudes (Arkes, 1991).

the location of one or more of the judge's response options. In the experiments, it is unclear what the precise nature of any clinical error will be, so it is unclear whether an incentive would make a difference based on Arkes' framework.

Given this uncertainty and the mixed evidence overall, the experiments employ a financial incentive. As indicated, several theories suggest that a financial incentive would increase average clinical accuracy. There is some (albeit a small) average effect across the meta-analyzed studies in Jenkins et al. (1998); in that meta-analysis, the numbers of meta-analyzed studies and subjects are small; there is an indication of moderators (credibility interval = 0 to 0); and the authors of that meta-analysis expressed concerns about the non-independence of coefficients. In addition, while it is possible that judgment errors in the experiments are associational or psychophysical, they may be strategic as well. It is not unreasonable to assume that, without incentives, strategy-based judgment errors due to idleness are a large component of error for undergraduate subjects. Although it might encourage association-based errors in judgment, there is no obvious reason why a financial incentive would do any harm (aside from decreasing available research funds).

RESEARCH PURPOSE AND QUESTIONS

The overall purpose of this research is to better understand clinical judgment and its prospects for accurate prediction. As mentioned at the beginning of this dissertation, there are six specific research questions, as follows:

1. To what extent (if any) can assessors learn to outperform a mechanical approach in the applicant selection context?
2. To what extent (if any) can assessors learn to improve their judgment validity (in terms of predictive accuracy)?
3. To what extent (if any) can assessors become less overconfident in their judgment strategies?
4. What is the relationship (if any) of any changes in predictive accuracy to any changes in overconfidence?
5. What individual differences (if any) define those individuals who predict, and learn to predict, most accurately?
6. When and to what extent (if any) do individuals possess insight about the optimal strategy for making accurate judgments?

Clinical Validity Versus Mechanical Validity

The first research question concerns the extent to which assessors can learn to outperform a mechanical approach in the applicant selection context. Therefore, one outcome of interest is the difference in criterion-related validity between clinical data combination and mechanical data combination after and during training. One dependent variable is the difference between judgment validity (r_a) and mechanical validity (R_e). Another dependent variable is the difference in skill scores between the mechanical and clinical approaches.

Change in Clinical Accuracy

The second research question concerns the extent to which assessors can learn to improve their judgment validity (in terms of predictive accuracy). Changes in r_a and clinical skill scores over time will capture the extent of that learning. Hierarchical linear modeling that treats r_a and skill score values across time points as the dependent variables, subjects as a grouping variable, and time points as nested within each subject can convey one measure of the extent to which learning occurred and whether it differed based upon the nature of feedback received (i.e., among different treatment conditions).

It is important to understand how and why clinical validity changes. Conclusions from simply looking at end products (r_a values and skill scores) are limited. By modeling change in the mechanisms underlying change in end products (i.e. use of unmodeled knowledge, use of mechanical knowledge, etc.), one can better understand what judges did or failed to do to improve their accuracy. As discussed, the decomposition of skill scores proposed by Murphy (1988), Stewart (1990), and Cooksey (1996) involves the addition of a measure of shape (r_{YO}) that excludes information provided by the *C* Lens Model parameter that is critical to understanding clinical accuracy (Cooksey, 1996). The *C* measure of unmodeled knowledge is what assesses the extent to which the judge uses the disordinal interaction (or any criterion-related information absent from the purely mechanical model) to make predictions. Although it assesses a between-studies relationship, it is worth mentioning that the meta-analytic correlation which Karelaia and Hogarth (2008) found between C and r_a was 0.23 ($p < 0.01$, $k = 204$ task environments), suggesting the possibility that within-person changes over time in r_a may be positively related to changes in use of unmodeled knowledge. There is also some evidence that suggests that C will increase in value as the gap between clinical and mechanical validity (r_a) closes.[53]

[53] Based on an analysis of the data at http://dx.doi.org/10.1037/0033-2909.134.3.404.supp underlying the Karelaia & Hogarth (2008) meta-analysis, the correlation between (a) improvement in clinical validity

Therefore, only a Lens Model decomposition will be used. Not only will this dissertation examine change in C over time, but it also will examine change in the criterion-related validity of unmodeled knowledge (r_z). The criterion-related validity of C is key to accuracy; C alone may contain criterion-irrelevant information. In Koele (1980), average r_z across 4 task ecologies increased from 0.17 to 0.36 after training. Based on data underlying Karelaia & Hogarth, 2008 (retrieved at http://dx.doi.org/10.1037/0033-2909.134.3.404.supp), average r_z increased after training (albeit from only 0.01 for 95 analyzable task ecologies to only 0.02 for 151 analyzable task ecologies).

Furthermore, change in G across time can indicate the extent to which change in judges' accuracy is related to change in the judge's use of knowledge of the mechanical process. Although maximization of mechanical knowledge alone might not allow the clinician to outperform the mechanical process, it nevertheless may contribute to accuracy. It has been demonstrated that values for G frequently are highly positively skewed and sensitive to the task environment (Castellan, 1992; Stewart, 1994). Nevertheless, the meta-analytic correlation between G and r_a was 0.78 ($p < 0.01$, $k = 236$ task environments; Karelaia & Hogarth, 2008), suggesting the possibility that improvement in predictive accuracy over time is related to increased use of mechanical knowledge.

In addition, change in cognitive control (R_s) will be measured to assess the assessors' abilities to use a judgment policy (i.e., a weighting and functional form approach) consistently over time. Meta-analytic data reported by Kareleia & Hogarth (2008) are inconsistent with conclusions drawn by Hammond, Hursch, & Todd (1964) and Einhorn (1974) that R_s and r_a are independent of each other. Based on the Lens Model Equation,

$r_a = GR_eR_s + C\sqrt{(1-R_e^2)}\sqrt{(1-R_s^2)}$, it seems that change in R_s could systematically affect r_a if

---

in comparison to mechanical validity and (b) C was 0.41 for the 95 learning task ecologies for which this correlation could be calculated (although almost all of these task ecologies are problematic in their constraint on the value of C and/or for some other reason already discussed) and 0.63 for just the 4 task ecologies in Koele (1980).

there are systematic trends for the other Lens Model components such that R_s and $\sqrt{(1-R_s^2)}$ do not offset each other in terms of their impact on the equation. In fact, the meta-analytic correlation between R_s and r_a is 0.56 ($p < 0.01$, $k = 237$ task environments; Karelaia & Hogarth, 2008). Evidence about the impact of training on the relationship between R_s and r_a is mixed.[54] It therefore would be useful to gain further evidence about the relationship between R_s and r_a.

As a way to concurrently measure the judge's reliance upon the disordinal interaction, change in the C_{xj} relative weight for the disordinal interaction will be assessed. Better judgment accuracy is indicated by more reliance (more weight) being placed upon the interaction term. This information is not mere redundancy. In theory, C might contain variability from the assessor's use of not just the disordinal interaction, but it also might contain variability from use of "broken legs" and other functional forms. Also, C might contain correlated error. Therefore, if change in the C_{xj} relative weight matches change in the C values, then there is greater confidence that change in C is due to greater reliance on the disordinal interaction feedback. Note that the C_{xj} relative weight for a functional form cannot replace C, because in most scenarios (particularly naturalistic settings), it is not definitively known in advance that use of that functional form will certainly improve judgment accuracy.

---

[54] Meta-analytic evidence indicates that any reduction in the judgment residual ($1 - R_s^2$) possibly due to the use of unmodeled cues or functional forms after the training was strongly associated with *even worse* clinical validity in comparison to mechanical validity. The correlation between how much the mechanical advantage over clinical data combination declined or reversed with change in R_s after training was -0.67 for 194 task environments (based on the data underlying Karelaia & Hogarth, 2008, retrieved at http://dx.doi.org/10.1037/0033-2909.134.3.404.supp). For 168 laboratory ecologies, the correlation was only slightly less unfavorable at -0.65. For 26 field studies, the correlation plummeted to -0.80. However, virtually all of the meta-analyzed studies differ substantially in experimental design from the studies being conducted in this dissertation (e.g., in terms of the availability of unmodeled knowledge, the predictibaility of the criterion, etc.). The correlation between how much the mechanical advantage over clinical data combination ($\sqrt{R_e^{\,2} - r_a^{\,2}}$) declined or reversed with change in R_s after training for the four tasks studied by Koele (1980) was a *positive* 0.36, suggesting that the experiments of this dissertation, which possess task characteristics comparable to those in Koele (1980), might similarly demonstrate a positive relationship between learning and the improvment of clinical data combination.

Change in Confidence

The third research question about the extent to which assessors become less overconfident in their judgment strategies also can be addressed with *d*-values, *t*-testing, and hierarchical linear modeling. For all of these methods, self-reported absolute and relative levels of confidence serve as dependent variables. The hierarchical linear modeling treats subjects as a grouping variable, and time points as nested within each subject can determine whether and how confidence changed and whether such change differed based upon the nature of feedback received (i.e., among different treatment conditions).

Relationship of Change in Clinical Accuracy to Change in Confidence

The fourth research question focuses on the relationship of any changes in predictive accuracy to any changes in overconfidence. The results from the second research question can be compared to the results of the third research question.

Individual Differences

Not only is it important to ascertain the extent to which people can be taught to be more accurate and less overconfident, but also it is vital to determine what kinds of people predict, and learn to predict, most acccurately. Poor judgment and training to improve judgment accuracy both have costs. The fifth research question, which asks about possible individual differences that define those individuals who (a) predict and (b) learn to predict most accurately, can employ correlation as well as analysis of regression models (whether they be hierarchical linear models or OLS models with fixed effects only). The correlation between measures of each subject's predictive accuracy before any feedback is received (r_a at time 1 and skill score at time 1) and individual difference variables will answer part (a) about people who presumably have not been trained by the methods used in the experiments. Correlations between individual differences and each measure of each subject's predictive accuracy after time 1 (i.e., at time 2, at time 3, etc.) will answer part (a) about people who have received varying amounts of training. Essentially, these analyses provide static snapshots. In order to answer part (b), about whether individual difference variables matter once change due to time is accounted for, analysis of regression models is needed. For each individual difference, the linear mixed effects models that are the most complex models that answer the second research question about change in clinical validity would be compared for fit against the same models with the individual difference added as a second-level variable.

The analyzed individual differences of main interest will include the following: (1) cognitive ability and achievement (general, verbal, mathematical, writing, reading comprehension, and scientific reasoning ability; college GPA; number of college academic hours completed; high school rank), (2) gender (male or female), (3) personality (extroversion/positive emotionality, neuroticism/negative emotionality, intellectance/conventionality, agreeableness, conscientiousness, positive valence, and negative valence), (4) interests (realistic, investigative, artistic, social, enterprising, and conventional), and (5) prior experience (number of outcomes for which a subject has consciously predicted human behavior prior to study participation, number of persons previously judged, period of time over which judgments were made, number of outcomes for which a subject has consciously predicted human behavior in a formal process, number of outcomes for which a subject has consciously predicted human behavior after the subject received training, length of any training, amount of formal statistics education, and amount of formal education in judgment and decision making). In Table 5 appears a list of these variables and scale descriptions. The forms being used to collect individual differences information appear in Appendices J through O.

Looking at these potential moderators makes sense based on theoretical considerations and/or empirical evidence. Clinicians higher in cognitive ability might possess better working memory (Kyllonen & Christal, 1990) and thus be able to avoid many of the cognitive pitfalls of clinical data combination enumerated earlier in this dissertation. Empirical evidence strongly suggests that higher general cognitive ability leads to greater training success ($r = 0.26$; Schmidt & Hunter, 1998). In theory, verbal, mathematical, writing, reading comprehension, and scientific reasoning ability all would be conducive to more accurate prediction, and it is possible that one specific ability will be more helpful than another. Cognitive predictors with a long-term motivational component (achievement predictors) such as high school rank and college GPA are expected to predict performance on the task worse than ability measures do, because performance on the task does not require such long-term commitment. Number of college academic hours earned might be predictive of task performance as a measure of how much university-level learning subjects have experienced. Meta-analytic evidence indicates that males tend to outperform females on some spatial cognitive tasks (Voyer, Voyer, & Bryden, 1995), and since the task information feedback is spatial in nature, it seems prudent to investigate if gender-based differences might partly explain between-subject differences in clinical performance.[55]

Given that the construct validity problems associated with self-reported grade point average and test scores (Kuncel, Credé, & Thomas, 2005), subjects were asked to consent to the release from their university of this information. The informed consent forms appear in Appendix J (Fall 2009) and Appendix K (Spring 2010). The Academic Information & Experience Form in Appendix L, collected self-reported cognitive ability information (in addition to experiential information) in case the subject did not consent to the university's release of academic information. Of the approximately 25% of participants for whom university-reported information was not provided, many were non-U.S. citizens who were unable to provide information for which norms were available (e.g., they provided a score for a test given only to secondary school graduates of their native country). It is possible that those who refused to provide consent to the university's release of their academic information were more likely to

[55] Thanks go to Winny Shen, a fellow Ph.D. candidate, for raising this issue.

inaccurately self-report their academic information. Therefore, only university-reported information for which norms were available was ultimately used in analyses.

Personality and interests are expected to moderate the relationship between predictive accuracy and time. Those who are more open to learning new ways to combine information or who are more diligent in task performance may outperform and learn more than those who are more careless or close-minded to new approaches. Among the strongest job performance relationships involving personality predictors are extroversion, openness, and conscientiousness predicting training proficiency ($\hat{\rho}$ = 0.26, 0.25, and 0.23, respectively; Barrick & Mount, 1991). The Goldberg (Goldberg, 1990) and IPC-7 (Tellegen, Waller, & Grove, 1987) inventories, which appear in Appendices M and N, respectively, were used to collect self-reported personality information from subjects. Based on Holland (1997)'s RIASEC model of career interests, those with investigative (experimental) interests "[u]se investigative beliefs, competencies, and values to solve problems at work and in other settings. . . . [, r]el[y] on careful analyses, objective data, and related scholarly practices [, and p]ay[] less attention to personal feelings and the social environment. . . . [H]e or she is apt to be analytical [and] precise. . . ." (p. 23). Motivation is a key determinant of successful performance (Campbell, 1990). Therefore, people with higher investigative (experimental) interests might be more motivated, and thus more successful, in scrutinizing the experimental judgment task, how best to perform it well, and how best to learn to improve performance. The RIASEC Interest Profiler Form (Rounds, Su, Lewis, & Rivkin, 2009), which appears in Appendix O, was used to collect self-reported information about subjects' interests.

Although differences in prior experience and the expectations that result from it (i.e., expertise) are not expected to result in any difference in judgment performance among subjects, this dissertation nevertheless examines this individual difference due to its popularity and the controversy surrounding it. Expertise is perhaps the most studied individual difference in judgment research, and the findings for its relationship to judgment accuracy seem highly counter-intuitive at first. "In tasks requiring decision under uncertainty, such as evaluating applicants for medical internships (Johnson, 1988) or predicting successes in graduate school (Dawes, 1971), it has been shown consistently that experts fail to make better judgments than novices" (Chi, 2006). Even after judges have gained years of experience, their judgment policies do not appear to grow more similar to each other (Brehmer & Brehmer, 1988).

Based on the meta-analytic findings reported by Karelaia & Hogarth (2008) for the effect of level of expertise on values of R_e (mechanical validity) and r_a (clinical validity), it appears that the superiority of mechanical over holistic predictive validity as measured by $\sqrt{R_e^{\ 2} - r_a^{\ 2}}$ is actually *largest* for experts (novices = .54 for 204 task environments; some training = .42 for 15 task environments; expert = .61 for 29 task environments). Using the same metric but based on the data underlying the learning studies for that meta-analysis (see http://dx.doi.org/10.1037/0033-2909.134.3.404.supp), it appears that training experts to make better judgments simply makes them just as bad as novices rather than worse (with the mechanical advantage over novice clinicians growing from a correlation increment of 0.45 to 0.50; over those clinicians with some prior training growing from 0.37 to 0.25; over expert clinicians growing from 0.31 to 0.49).[56] It has been argued that these types of pessimistic findings for expertise are limited to the prediction of human behavior (Shanteau, 1984). In any case, self-reported information about experience and expertise in predicting human behavior was collected using the Academic Information & Experience Form which appears in <u>Appendix L</u>.

These findings regarding expertise are consistent with previous discussion. Why should one expect experience to be an effective teacher given the long list of limitations on learning, biases, faulty heuristics and other obstacles that beset the clinician in predicting human behavior?

[56] One possible implication is that cognitive task analysis (CTA), which many NDM researchers employ to investigate the strategies that experts use (see Crandall, Klein, & Hoffman, 2006), might not be useful in improving the prediction of psychological outcomes.

One need only reconsider what already has been discussed in the context of why clinical data combination would be expected to underperform mechanical data combination. Real world tasks entail unavoidable uncertainty, and people typically fail to understand and accept this reality. Even with feedback (which people might reject or fail to comprehend), they draw incorrect conclusions about success and failure, how to achieve the former, and how to avoid the latter – and real world judgment often involves no helpful feedback at all. The feedback that is received usually is limited to outcome feedback, and as discussed, this type of feedback is often useless and counterproductive. Even when useful feedback is provided and accepted, the clinician still must succeed on several fronts: identify which cues are relevant and which are not, identify the functional forms of the relevant cues (particularly when the task ecology is non-monotone), find proper weights to give the cues (particularly when unit weighting is expected to substantially underperform optimal weighting), and identify the inter-correlations among the cues (Brehmer, 1979; Klayman, 1988). What success people do experience may be largely due to the generosity inherent to judgment tasks, namely the robustness of linear models and unit weighting and the existence of redundancy of cue information, which increases judges' opportunity to acquire at least some accurate information about the task environment (Klayman, 1988).

Insight

The sixth question concerns when insight[57] is acquired about the optimal strategy for making accurate judgments as well as the extent of that insight. Identifying insight is challenging. It is at best difficult to know the true character of the assessor's judgment approach.

[57] The Oxford English Dictionary (http://dictionary.oed.com, 2010) defines insight in the following ways:

- "In studies of behaviour and learning, the sudden perception of the solution to a problem or difficulty; applied to animals, giving an indication of their capacity for ideas and reasoning."
- "Internal sight, mental vision or perception, discernment; in early use sometimes, Understanding, intelligence, wisdom."
- "The fact of penetrating with the eyes of the understanding into the inner character or hidden nature of things; a glimpse or view beneath the surface; the faculty or power of thus seeing."

It is internal to the person and involves more than just behavior, if it involves behavior at all. As discussed, any model of judgment is a “paramorphic representation” (Hoffman, 1960). Brehmer & Brehmer (1988) describe how infrequently assumed judgment policy models are tested for their validity, making impossible any comparison of these models against narrative self-reports of judgment policies. Furthermore, high judgment accuracy might be due to chance or the reflexive application of a recommended methodology (i.e., just following orders). Therefore, simply measuring performance with Lens Model parameters or skill scores is helpful but insufficient for assessing insight. For an outsider (and perhaps even for the judge himself) to accurately understand whether insight is achieved, several obstacles must be surmounted. First, judges need to truly understand the judgment processes that they use. Next, judges need to be honest about whether they have achieved insight and the nature of that insight. Finally, judges need to be able to articulate with sufficient clarity that they have attained insight and the nature of that insight. If any of these three hurdles is not overcome, then it is at best difficult for a third party (and perhaps the judge himself) to know if, when, and how insight is attained. As Brehmer and Brehmer (1988) discuss, these goals are difficult to achieve, and at least as of their literature review, little is known about how best to achieve these goals.

In order to address these concerns, judges were directly asked after making each of several sets of predictions if they had used the disordinal interaction. Since judges sometimes had not been explicitly told anything about disordinal interactions at that point in their experimental condition, judges sometimes instead were asked to describe the relationship between cognitive ability and performance. However, a judge’s answer to either question might be inaccurate, because the judge might misunderstand what the question was asking or the nature of the judge’s strategy, or the judge might be less than candid. Therefore, prior to being asked about the disordinal interaction or the relationship between cognitive ability and performance, judges were asked to generally describe the prediction strategies that they had used. After being asked if they had used the disordinal interaction, judges were then prompted to specifically describe how they had used the disordinal interaction.

Furthermore, it is possible that a judge had insight about the existence of the disordinal interaction (i.e., understood that the relationship between cognitive ability and performance was positive if the job was interesting but negative if the job was boring) but lacked insight about how to actually use this information to make a judgment. Therefore, responses were coded on the following 3 dimensions: (1) whether the subject said that s/he had used the disordinal interaction, (2) whether the subject expressed awareness of the disordinal interaction (i.e., that the relationship between cognitive ability and performance depended upon the interesting or boring nature of the job), and (3) whether the subjects used the disordinal interaction properly based on the subject's narrative description of his/her strategy. Due to the vagueness of most subjects' narrative responses, some responses were not codeable, and in general coding was binary in nature (i.e., 1 for "yes", 0 for "no"). This approach seems sensible given findings that attempts to model verbal descriptions of judgment policies usually lead to over-complexity (Brehmer & Brehmer, 1988). Coding for all dimensions for all subjects was conducted separately by two Ph.D. candidates and later reconciled after discussion.

It is unclear whether or not to consider an assessor to have gained insight if that assessor experienced a "reversal." An example of a "reversal" is when coding indicates that a judge previously had expressed awareness of the disordinal interaction and then for a later block the coding indicates that the judge did not express awareness of the disordinal interaction. Reasons for the reversal could be due to a true failure of the judge to have insight, a failure of a judge with insight to be sufficiently articulate when writing responses for just one or two particular blocks, a coding error, or some other reason. One of the main purposes of the experiments is to determine whether assessors can learn to *systematically* outperform a mechanical approach. It can plausibly be argued that *any* reversal creates doubt about the ability of a judge to meet that performance standard. Of course, it is always possible that (1) if an additional block of judgments had been provided, judges who previously experienced no reversals would have experienced one for an added block and (2) that if fewer blocks had been provided, no reversals would have been detected for some judges who experienced a reversal for one of the later blocks.

One hypothesis is that differences in the sample-level feedback from block to block create occasional uncertainty even among assessors with insight about the importance of the disordinal interaction. Assuming that a person is able to mentally separate out the levels of the moderator (interesting versus boring), the disordinal interaction would be clear based on the population of data on which it is based but not as clear for a sample of that population of data. One of the keys to insight is the understanding of sampling error. To gain insight, an assessor needs to understand the functional form not just for a sample in isolation but instead as existing for a population for which noisy, sample-level appear from time to time. In other words, the judge must appreciate trends across samples that may not hold for each and every sample while simultaneously developing a judgment strategy that ignores data patterns that uniquely exist for particularly noisy samples. Ideally, clinical assessors would cumulate data to better capture population-level patterns. Given the human cognitive limitations discussed throughout this dissertation, this hope seems unrealistic especially when the amount of data is not small.

Since reversals raise fundamental issues about insight, they were reported. As there are several possible ways of measuring insight when reversals exist, this dissertation interpreted the existence of reversal in several different ways (see Table 16 and Table 25). Further research is needed to determine how many blocks of consistent responses would be needed, on average, to indicate that true insight has been acquired.

METHODS

Differences Between Fall 2009 and Spring 2010 Designs

The experimental portions of the Fall 2009 and Spring 2010 studies (sometimes collectively referred to as "the experiments") differed in several ways (see Appendix A for a list and description of major differences in design). One noticeable difference is in the number of measured time points for Fall 2009 (4) and Spring 2010 (5). For the Fall 2009 dataset, the fifth block of judgments was originally intended to measure temporal stability of judgment by incorporating profiles from each of the prior blocks. However, the goal of the experiments is learning (the opposite of temporal stability), so it was later decided that measurement of temporal stability would not be fruitful. Furthermore, the first four blocks of the Fall 2009 dataset were selectively chosen from a larger pool of simulated data (sometimes referred to as the "population") so that their task information characteristics (cue intercorrelations, cue-criterion relationships, etc.) were deemed interpretable in graphical feedback format. This action was undertaken to reduce the noise due to sampling error common to samples of data from populations of data. Given that the population functional form (generated from 2,000 profiles) was a disordinal interaction, a randomly chosen block of 25 cases from the population would probably have resulted in uninterpretable task information feedback. As a sampling from prior blocks, the fifth block of the Fall 2009 dataset suffered from this lack of interpretability. Given the didactic goal of the experiments and the difference in interpretability between the fifth block and the prior blocks, it was decided to ignore this fifth block for the Fall 2009 dataset. Since the Spring 2010 dataset did not attempt to assess temporal stability or selectively choose blocks for interpretability, the fifth block for the Spring 2010 dataset was retained.

The experimental design changed between Fall 2009 and Spring 2010 partly in the hope that a new experimental blueprint would permit a clearer determination of the reasons why clinicians generally fail to outperform mechanical approaches. It was believed that changes with regard to feedback which are reflected in Appendix A would facilitate better discrimination between those subjects who could "see" (i.e. conceptualize) the interaction but not effectively employ it and those who could not even "see" it. To help achieve that result, the disordinal interaction was stronger for the Spring 2010 study than for the Fall 2009 study. It was hoped that a stronger disordinal interaction would (a) be more salient to subjects and (b) by their stronger nature would permit subjects to outperform a linear version of a non-linear mechanical model that included the disordinal interaction. The strength of the disordinal interaction for the presented data, the strength of the disordinal interaction for the populations on which those data were based, and the strength of the linear mechanical models (R_e) appear in Tables 3 and 4. Both at the sample and population level, the disordinal interaction is stronger in the Spring 2010 dataset as measured by R^2 and the C_{xj} relative weight. Based on absolute ΔR^2 from adding the disordinal interaction to a linear model, the disordinal interaction in the Spring 2010 study is almost always stronger at both the sample and population levels. Based on relative (%) ΔR^2 from adding the disordinal interaction to the linear model, the Fall 2009 disordinal interaction is stronger at both the sample and population level. It is expected that disordinal interaction that is stronger is easier to conceptualize. If one measures that strength based on R^2, absolute ΔR^2, or the C_{xj} relative weight, one would expect (all other things being equal) that subjects would perform better in Spring 2010. If one measures that strength based on relative (%) ΔR^2, one would expect (all other things being equal) that subjects would perform better in Fall 2009. Again, the majority of measures indicate that the Spring 2010 disordinal interaction is stronger and therefore more likely to be recognized and successfully used by the clinician to outperform the linear mechanical model.

Procedures

Two separate survey-based studies that employed similar designs were undertaken in the Fall of 2009 and the Spring of 2010 (sometimes collectively referred to as "the studies"). Data and graphical feedback used for these studies were generated using R, SPSS, and Microsoft Excel software. Each subject in both studies was given 2 hours to complete the study in paper and pencil format in a proctored setting. Prior to the experimental portions of the studies, subjects were asked to respond to surveys that collected information about individual differences (cognitive ability and achievement, gender, personality, interests, and experience). Appendices L through O contains copies of the surveys used in both studies. In addition to consenting to their participation in the studies, subjects were specifically asked to consent to the release of their academic records as part of the process of collecting cognitive ability and achievement information. Copies of the consent forms appear in Appendices J and K.

The experimental portions of the Fall 2009 and Spring 2010 studies followed the collection of individual differences information. They also both followed and preceded the collection of information from subjects about subjects' absolute and relative levels of confidence[58] in their ability to make accurate predictions in the studies overall. The wording and format of the questions about confidence that preceded and followed the experimental portion of the studies appear in the Academic Information & Experience Form in Appendix L and the first two questions in Appendix Q, respectively. In both experiments, subjects were asked to make predictions of job performance for hypothetical job applicants based on scores of 2 positively correlated cues that were provided to subjects for each and every applicant. (Each set of scores for an applicant is sometimes referred to as a "profile".) These predictions were made for each of 125 hypothetical job applicants who were divided into groups of 25 subjects judged over 5 blocks. For the Fall 2009 study, the last (5^{th}) block of 25 judgments was eliminated from later analyses, so the Fall 2009 study is usually described and treated as if only the first 4 blocks of 100

[58] "Relative confidence" is defined as the confidence one has in one's performance relative to other participants. "Absolute confidence" is defined as confidence that is not intended to reflect beliefs about others' performance.

predictions were made.[59] The 2 cues consisted of each applicant's score on a cognitive ability test as well as whether or not the applicant was expected to find the job for which the applicant was applying interesting or boring (see Appendix C for Fall 2009 and Appendix G for Spring 2010).[60] After each block of judgments was made, subjects were asked to self-report their absolute and relative levels of confidence in how accurate they were in making predictions for that block (see Appendix P). They then were asked to describe their judgment strategies used for the prior block, including whether and how they used the disordinal interaction (see Appendix P).

Appearing next was graphical feedback about the statistical relationships among the cues and between the cues and the job performance criterion for the immediately prior block of predictions and/or for the population of every applicant who would ever apply for the job(s).[61] The nature of this feedback varied by study (i.e., for Fall 2009 versus Spring 2010) as well as by experimental condition within each study (see Appendix E for Fall 2009 task information feedback and Appendix I for Spring 2010 task information feedback). Each study contained 3 experimental conditions, each of which provided a different amount and different types of

[59] The Fall 2009 study (but not the Spring 2010 study) later eliminated the data for the last 25 predictions (leaving 4 blocks of 100 judgments), because those predictions were originally designed to measure temporal stability of predictions. For that purpose, the last (5th) block consisted of 25 profiles sampled from earlier blocks. Unlike the first 4 blocks, which were specially selected to reduce sampling error (i.e., noise in the data), the last block contained very noisy data. The data for the 5th block were eliminated after all subjects participated, because the purpose of the Fall 2009 and Spring 2010 studies was actually to measure instability (i.e., learning), and the noisier nature of the data from the 5th block made it incomparable to the data for the earlier blocks.

[60] The cues and criterion differed between the 2 studies in how they were scaled. In the Fall 2009 study, cognitive ability and how interesting or boring the job would be to the applicant were measured on a 0 to 6 anchored Likert scale. Subjects were asked to make predictions of job performance on a 0 to 6 anchored Likert scale, too. These scales appear in Appendix B and Appendix C. In the Spring 2010 study, cognitive ability was measured on a percentile scale. Subjects were asked to make predictions of job performance based on a percentile scale, too. How interesting or boring the job would be to each applicant was provided on a dichotomous scale ("Interesting" or "Boring"). These scales are depicted in Appendix G. It was thought that understanding and use of these scales would be more intuitive to subjects than the 7-point Likert scales used in the Fall 2009 design.

[61] In the Fall 2009 study, hypothetical applicants were not necessarily applying for the same job. As Hunter and Hunter (1984) describe, the relationship between cognitive ability and job performance may vary (at least somewhat in strength) for jobs of different complexity. To address the possibility that the statistical relationship between the cognitive ability cue and job performance would vary across different jobs, the Spring 2010 study explicitly described applicants as applying for the same job.

graphical feedback about cue and criterion statistical relationships underlying the judgment task (task information feedback, discussed in more detail below). For both experiments (albeit in not exactly the same way), feedback condition 1 was designed to provide the least amount of feedback, particularly about the disordinal interaction and whether and how to employ it; feedback condition 2 was designed to provide a moderate amount of feedback information, particularly about the disordinal interaction and whether and how to employ it; and feedback condition 3 was designed to provide the greatest amount of feedback information, particularly about the disordinal interaction and whether and how to employ it. Appendix D describes more specifically the major differences among the 3 feedback conditions for the Fall 2009 experiment, and Appendix H describes more specifically the major differences among the 3 feedback conditions for the Spring 2010 experiment. After this feedback information appeared, subjects would encounter a new block of 25 predictions to be made, followed by another round of questions about confidence and judgment strategies and then graphical feedback about the statistical nature of that new block. This process repeated itself until predictions and self-reports of confidence and prediction strategies were made for all blocks.

Subjects

Subjects consisted almost solely of undergraduates at the University of Minnesota. Almost all of them were enrolled in at least one psychology course, and many of them were enrolled in introductory psychology. Subjects knew in advance that as compensation for making a good faith effort in the study, they would receive extra credit points that could be used toward their grade in a class. They also knew before participating that those subjects who predicted most accurately in their respective experimental condition would receive $75, second-most accurately in their experimental condition $50, and third-most accurately in their experimental condition $25. For the Fall 2009 study, there were 181 participants. Only 142 of these participants provided analyzable data (48 in feedback condition 1, 43 in feedback condition 2, and 51 in feedback condition 3). For the Spring 2010 study, there were 159 participants. Only 141 of these participants provided analyzable data (46 in feedback condition 1, 48 in feedback condition 2, and 47 in feedback condition 3). Reasons for exclusion of subjects' data included failure to respond to a large percentage of items, signs of random responding, and indications that subjects' lost their place on the scantron forms which they were using to respond to survey items.

RESULTS AND DISCUSSION

Clinical Validity Versus Mechanical Validity

The first research question concerns the extent (if any) to which can assessors learn to outperform a mechanical approach in the applicant selection context. Tables 6 and 17 summarize the narrative explanation of the statistical analyses below for Fall 2009 and Spring 2010, respectively. In Figures 10 through 13 and Figures 46 through 49 appear bar and line graphs illustrating mechanical and clinical accuracy measured at the end of and during the experiments from Fall 2009 and Spring 2010, respectively.

Fall 2009 Results

Based on comparisons of experimental treatment group means of r_a and skill score to the mechanical linear model's criterion related validity (R_e) and skill score based on the population, it appears that for subjects exposed to feedback condition 2, holistic judgment *is* more accurate than actuarial prediction (see Figures 10 and 11). Linear mechanical R_e equaled 0.168 and exceeded mean judgment validities (r_a values) of 0.118 for feedback condition 1 and 0.154 for feedback condition 3. However, it was smaller than the mean r_a value for feedback condition 2, which equaled 0.173. Linear mechanical skill score equaled 0.0102 and exceeded mean judgment skills scores of -1.841 for feedback condition 1, -1.464 for feedback condition 2, and -1.29 for feedback condition 3. If one limits the measure of accuracy to r_a (similarity in rank order or shape) only (i.e., if one ignores elevation and scatter information of the skill score), then one can say that trained subjects were able to outperform the linear actuarial model.

Are these results statistically significant? It would appear that one should *not* conclude that any judges whomsoever trained in feedback condition 2 can outperform the mechanical method. One reaches this conclusion after calculating a confidence interval for the r_a value of feedback condition 2 (0.173) as well as testing the statistical significance of the difference between the r_a for feedback condition 2 and the linear mechanical validity R_e. The 95% confidence interval is [-0.134, 0.449] (N = 43) and clearly contains the linear mechanical R_e value of 0.168. For the significance test, one may use the equation $z = \frac{Z_r - Z_p}{\sqrt{\frac{1}{N-3}}}$ to obtain an observed test statistic, where Z_r equals the Fisher r-to-z transformed r_a value for feedback condition 2, Z_p equals the Fisher r-to-z transformed R_e value for the linear mechanical approach, and N equals the 43 subjects in feedback condition 2 (Cohen, 2001, p. 268). This approach necessitates treating the linear mechanical validity R_e as if it represented the value for a theoretical population of purely mechanical judges to which the sample represented by the feedback condition 2 r_a value might belong. Based on the equation, z = 0.316. Comparing this value to a critical two-tailed z-value of 1.96 for an α-level of 0.05, one would fail to reject the null hypothesis that feedback 2 validity is greater than the validity of the linear actuarial model.

If the training had terminated sooner, results would still be statistically insignificant. As <u>Figures 12 and 13</u> indicate, the only other time when clinical accuracy exceeds mechanical accuracy is at Time 0 (for Block 1), before any feedback training is received. Whether additional blocks of training would have led to clinical accuracy that exceeds mechanical accuracy after accounting for the risk of sampling error is an interesting question that requires additional research. <u>Figures 12 and 13</u> suggest that this outcome would occur if linear trends as of Block 3 (Time 2) were to persist long enough. Of course, the functional form (growth curve) for changes in judgment accuracy during any type of training may be non-linear during any time period.

Spring 2010 Results

Based on comparisons of experimental treatment group means of r_a and skill score to the mechanical linear model's criterion related validity (R_e) and skill score based on the population, it appears that with regard to the subjects in the experiment and after the training experienced in all feedback conditions, holistic judgment measured by r_a but not skill score *is* more accurate than actuarial prediction (see Figures 46 and 47). Linear mechanical R_e equaled 0.367 and was smaller than the mean r_a values for the feedback conditions, which respectively equaled 0.433, 0.417, and 0.527. Linear mechanical skill score equaled 0.134 and exceeded mean judgment skills scores of -0.364 for feedback condition 1, -0.465 for feedback condition 2, and -0.244 for feedback condition 3. If one limits the measure of accuracy to r_a (similarity in rank order or shape) only (i.e., if one ignores elevation and scatter information of the skill score), then one can say that trained subjects were able to outperform the linear actuarial model.

Are these results statistically significant? One should *not* conclude that any judges whomsoever trained in any of the feedback conditions can outperform the mechanical method. One reaches this conclusion after calculating a confidence interval for each of the r_a values of the feedback conditions. For condition 1 ($r_a = 0.433$; $N = 46$), the 95% confidence interval equals [0.164, 0.642]. For condition 2 ($r_a = 0.417$; $N = 48$), the 95% confidence interval equals [0.151, 0.626]. For condition 3 ($r_a = 0.527$; $N = 47$), the 95% confidence interval equals [0.283, 0.707]. All of these intervals contain the linear mechanical R_e value of 0.367. Testing the statistical significance of the difference between the r_a for each feedback condition and the linear mechanical validity R_e (0.367) leads to the same conclusions. The z significance test here is the same as the one used for analyzing the same research question for the Fall 2009 data (Cohen, 2001, p. 268). This approach necessitates treating the linear mechanical validity R_e as if it represented the value for a theoretical population of purely mechanical judges to which the samples represented by the r_a values might belong. Based on the equation, $z = 0.54$ for condition 1, 0.41 for condition 2, and 1.38 for condition 3. Comparing each of these values to a critical two-tailed z-value of 1.96 for an α-level of 0.05, one would fail to reject the null hypothesis for each feedback condition that the validity for that feedback condition is greater than the validity of the linear actuarial model.

If the training had terminated sooner, results would still be statistically insignificant. As Figures 48 and 49 indicate, clinical accuracy that exceeds linear mechanical accuracy is never larger than an r_a value of 0.527 (for condition 3). R_e is a constant 0.367. The sample sizes among the feedback conditions range only from 46 to 48. Since the confidence interval and significance test for an r_a value of 0.527 do not permit rejection of null hypotheses, then one does not obtain staitistically significant results for any time point. Once again, whether additional blocks of training would have led to clinical accuracy that exceeds mechanical accuracy after accounting for the risk of sampling error is an interesting question that requires additional research. Figures 48 and 49 suggest that this outcome would occur if trends for feedback condition 3 were to persist long enough. Of course, the functional form (growth curve) for changes in judgment accuracy during any type of training may be non-monotone (i.e., direction-reversing).

Discussion

For both the Fall 2009 and Spring 2010 datasets, there is some evidence that trained subjects can outperform a mechanical linear model when performance accuracy is measured by r_a (but not skill score). For the Fall 2009 study, this outperformance occurred when subjects received explicit information that the disordinal interaction existed (feedback condition 2). For the Spring 2010 study, this outperformance occurred in all feedback conditions. It is possible that the provision in all Spring 2010 feedback conditions of adjacent scatterplots showing the relationship between cognitive ability and performance separately for interesting jobs and then boring jobs boosted performance across all feedback conditions. Also, the greater strength of the disordinal interaction in the Spring 2010 might have contributed to the better clinical results for Spring 2010. Of course, other explanations may exist. Holistic outperformance of the mechanical approach for the Spring 2010 study was most pronounced when subjects were explicitly told that the disordinal interaction existed (feedback condition 3). The collective findings across studies suggest that across different task environments, more explicit feedback about the disordinal interaction is conducive to more improvement in holistic accuracy, and with certain task environment characteristics and feedback this facilitative effect can permit the judge to predict more accurately than the actuarial model.

These findings are problematic for advocates of the superior accuracy of clinical prediction. Assuming that there exist nonlinear functional forms that would permit judges to outperform a realistic mechanical model, judges generally are not told about the existence of nonlinear functional forms. They would have to find them by themselves. Furthermore, in the real world there is no researcher who lays out in graphical form the nature of any nonlinear functional form. Judges must come to understand such a form by themselves. In addition, interactions or other nonlinear components of a functional form might be weaker than the disordinal interactions provided in the experiments. In the Fall 2009 study, presented data were specially selected to minimize the noise (sampling error) and thus make the disordinal interaction more salient to subjects. Moreover, the findings of superior clinical accuracy are limited to, among other things, the measurement of predictive accuracy as r_a. In other words, the evidence indicates that if subjects are able to outperform the mechanical model, then they are able to do so only to the extent of predicting rank order. Although that may be adequate for some selection purposes, it should be noted that when elevation and scatter are part of the measurement of predictive accuracy (as they are in skill score) subjects never outperform the actuarial approach. Furthermore, superior performance of the judge was never statistically significant. Of course, the proponents of superior clinical accuracy in naturalistic settings might argue that with additional training findings would have been statistically significant. However, this is speculative and does not address concerns that real world situations do not advantage assessors like the experiments did. In conclusion, the evidence fails to show that training even under ideal conditions will empower judges to predict an outcome like job performance more accurately than would a relatively straightforward mechanical approach.

Change in Clinical Accuracy

The second research question concerns the extent (if any) to which assessors can learn to improve their judgment validity (in terms of predictive accuracy). In Fall 2009, group 3 received the earliest explicit feedback about the disordinal interaction, group 2 the second earliest explicit feedback about it, and group 1 no explicit feedback at all about it. In Spring 2010, group 3 received explicit wording that the disordinal interaction existed, group 2 explicit wording that it might exist, and group 3 no such explicit wording about its possible or actual existence. If past research indicates that criterion-relevant feedback would improve accuracy, then for both Fall 2009 and Spring 2010 one would expect group 3 to experience the most positive growth in r_a and skill score across time, group 2 the second most positive growth, and group 1 negative growth or the least positive growth.

One can look at sample-level data (e.g., Figures 12, 13, 48, and 49) to observe mean level changes over time. However, these findings do not address the risk of error in predicting future performance of people exposed to the same training environment. If one wishes to most accurately account for the risk of sampling error in estimating the direction and magnitude of a slope trajectory that measures change over time, then one would employ hierarchical linear modeling (HLM, sometimes referred to as multilevel linear modeling, linear mixed effects modeling, random effects modeling, random coefficient regression, and covariance components modeling). Therefore, future analysis and discussion of measures of judgment accuracy (r_a and skill score) and measures of the determinants of that accuracy (C, r_z, G, R_s, and C_{xj}) will be based upon HLM methodology.

When used for longitudinal designs, HLM is sometimes called growth curve modeling or simply applied longitudinal analysis. When its use is justified for longitudinal designs, HLM is employed to model grow curves (how a dependent variable changes over time). HLM is a form of regression that is appropriate when observations are clustered within groups. When such clustering occurs, the assumption in regression of independence of observations and independence of error terms is violated. Consequently, standard errors for regression weights are larger than they would be if the observations were independent and error terms were independent. In longitudinal studies, such as these, the responses (outcomes, dependent variable values) of each subject are clustered within each subject. In other words, one would expect a person's response at Time 1 to be more like that person's responses at Times 2, 3, 4, etc. than they would be like another person's responses. Consequently, HLM is often appropriate for a longitudinal research design (Fitzmaurice, Laird, & Ware, 2004). When it is not appropriate, theoretical and statistical parsimony (including preservation of degrees of freedom) dictate use of a fully fixed effects (i.e., traditional OLS regression) model.

Several steps exist for HLM analysis. Use of a form of the intra-class correlation coefficient (sometimes called the ICC(1), ICC-1, ICC(1,1), or simply ICC) is typical for detecting whether sufficient clustering has occurred to justify use of an HLM approach. It is equal to $\frac{MSB - MSW}{MSB + [(k-1) * MSW]}$, where MSB is the variability across the groups (clusters), MSW is the variability within the groups, and k is the group size (Bliese, 2000). It indexes interrater reliability (Bliese, 2000; Bryk & Raudenbush, 2002). In a longitudinal context, the ICC indicates the extent to which variability in the dependent variable across time is attributable to differences among people. When the ICC is large enough (> 0.05 for purposes of this dissertation), the model for the data will be assumed to have random intercepts (i.e., unlike in a traditional regression model that has only a fixed intercept, the intercept may differ across the individuals). The software program R also will permit a likelihood ratio test of a fixed effects model against a random intercepts model to determine if a random intercepts model fits the data better. However, unlike the measurement of the ICC, use of that test is not generally prescribed. (On the other hand, there do not appear to be any authorities who disapprove of its use either.) Also, even if the test shows a significantly better fit for the random intercepts model, a very small ICC value indicates that the difference among individuals in intercepts is small. Therefore, the likelihood ratio test will be used only to bolster the finding from the ICC analysis. Trellis plots can be created to try to "eyeball" differences in intercepts across people, too. Next, a model with a random slope (in addition to a random intercept) may be fit to the data based on the results of a log likelihood chi-square test. This, too, differs from the traditional regression approach in which the model slope is fixed. Trellis plots can be created to try to "eyeball" differences in slopes across people. Essentially, once a model is tested for a random slope, the HLM aspect of model fitting ends. Nevertheless, if there are experimental treatments (as here), then the model can also be tested for whether intercepts and/or slopes differ across the experimental treatments. Finally, the growth trajectories (slopes) for change in dependent variables over time can be measured. The slope for the fixed effect (slope average across all persons) is determined. If the addition of treatment conditions significantly improved model fit, then separate slopes for each treatment group can be measured. For a more detailed discussion of HLM in general, one can refer to Raudenbush and Bryk (2002). For a more detailed discussion of longitudinal research designs that use HLM, one can refer to Fitzmaurice, Laird, and Ware (2004). In Table 7 (Fall 2009) and Table 18 (Spring 2010) appear summaries of the statistical results from following the aforementioned steps to assess changes across time in the following dependent variables (all of which measure accuracy or determinants of accuracy): r_a, skill score, C, r_z, G, R_s, and C_{xj}. (For a reminder of what these variables represent, one may refer to Figure 1, Table 1, Table 5, and previous discussion.)

Fall 2009 Results.

Judgment criterion-related validity (r_a). The ICC for this dependent variable equals 0.27, indicating that the person accounts for 27% of the variability in r_a. The software program

R, which permits a likelihood ratio comparison of a fully fixed effects model with a random intercepts model, indicates that the random intercepts model is a better fit to the data ($\chi^2(1) = 50.086$, $p < 0.0001$). Therefore, a multilevel model (i.e., at least a random intercepts model) fits the data better than a fully fixed effects regression mode does. Furthermore, a random slopes (and intercepts) model fits the data better than does a random intercepts model based on a log likelihood chi-square test ($\chi^2(2) = 21.572$, $p = 0.00002$). Therefore, both intercepts and slopes vary across subjects in terms of their change in judgment accuracy as measured by r_a (i.e., rank-order or shape similarity between clinically predicted and observed job performance). The trellis plot in Figure 14 does suggest variation in intercepts and slopes across subjects. However, models that test whether random intercepts and random slopes vary by feedback condition did not improve fit of a random effects model ($\chi^2(4) = 5.228$, $p = 0.265$ for a random intercepts model; $\chi^2(4) = 3.941$, $p = 0.414$ for a random intercepts and random slopes model). Therefore, in terms of r_a as the dependent variable, differences among feedback conditions did *not* matter. Although the boxplots in Figure 15 indicate that intercepts and slopes do vary among feedback conditions, apparently this variation was not statistically significant. In Figures 16 and 17, respectively, appear fitted growth curves (regression lines) for all subjects as well as a single fitted growth curve for a fully fixed effects model (the average for subjects overall). The first graph is somewhat ambiguous as to whether the subjects tend to decline in accuracy over time, although some of the negatively sloped lines are especially steep in comparison to the positively sloped lines generally. As one can see, the fixed effect regression line is negatively sloped ($\gamma = -0.006$, but its 95% confidence interval of [-0.017, 0.005] contains 0 and positive values). To the extent that one can speak meaningfully about an overall growth curve trajectory, individuals tend to decline in accuracy as measured by r_a.

Skill score. The ICC for this dependent variable equals 0.35, indicating that the person accounts for 35% of the variability in skill score. The software program R, which permits a likelihood ratio comparison of a fully fixed effects model with a random intercepts model, indicates that the random intercepts model is a better fit to the data ($\chi^2(1) = 80.814$, $p < 0.0001$). Therefore, a multilevel model (i.e., at least a random intercepts model) fits the data better than a fully fixed effects regression model does. Furthermore, a random slopes (and intercepts) model fits the data better than does a random intercepts model based on a log likelihood chi-square test ($\chi^2(2) = 7.912$, $p = 0.0191$). Therefore, both intercepts and slopes vary across subjects in terms of their change in judgment accuracy as measured by skill score (elevation and scatter in addition to rank-order or shape similarity between clinically predicted and observed job performance). The trellis plot in Figure 18 does suggest variation in intercepts and slopes across subjects. Furthermore, a model that tests whether random intercepts and random slopes vary by feedback condition improves fit of a random effects model ($\chi^2(4) = 11.739$, $p = 0.0194$ for a random intercepts and slopes model). Therefore, in terms of skill score as the dependent variable, differences among feedback conditions mattered. The boxplots in Figure 19 show how intercepts and slopes vary among feedback conditions. In Figure 20 appear fitted growth curves for each of the feedback conditions. The more feedback a group received about the disordinal interaction, the better its growth trajectory ($\gamma_{group1} = -0.134$ with a 95% confidence interval of [-0.254, -0.014], $\gamma_{group2} = 0.032$ with a 95% confidence interval of [-0.095, 0.158], and $\gamma_{group3} = 0.132$ with a 95% confidence interval of [0.016, 0.248]). In fact, the group that received the least feedback (i.e., group 1, which was never explicitly told about the existence of the disordinal interaction) experienced negative growth. There is a statistically significant difference in trajectory between groups 1 and 3. It may be possible that if sufficiently trained under feedback condition 3, a judge would eventually outperform a linear mechanical model. Based on the overall fitted growth curve, the average change in skill score overall is positive but not statistically significant ($\gamma = 0.012$ with a 95% confidence interval of [-0.074, 0.097]). These findings are consistent with the idea that increased feedback about the disordinal interaction would lead to improved predictive accuracy and that lack of such feedback leads to poorer predictive accuracy.

Unmodeled knowledge (*C*). The ICC for this dependent variable equals 0.23, indicating that the person accounts for 23% of the variability in *C*. The software program R, which permits a likelihood ratio comparison of a fully fixed effects model with a random intercepts model, indicates that the random intercepts model is a better fit to the data ($\chi^2(1) = 37.727$, $p < 0.0001$). Therefore, a multilevel model (i.e., at least a random intercepts model) fits the data better than a fully fixed effects regression model does. Based on an alpha-level of 0.05, random slopes (and intercepts) model barely fails to fit the data better than does a random intercepts model based on a log likelihood chi-square test ($\chi^2(2) = 5.791$, $p = 0.055$). However, one should remember that the choice of using an alpha-level of 0.05 is merely historical convention (Cowles & Davis, 1982). If one adopts an alpha level of 0.06, then there is statistically significant variation across individuals in both intercepts and slopes of the growth curves for use of unmodeled knowledge (presumably the disordinal interaction). The trellis plot in Figure 21 does suggest variation in intercepts and slopes across subjects. Furthermore, a model that tests whether random slopes and intercepts vary by feedback condition using an alpha level of 0.05 barely fails to improve fit of a random intercepts and slopes model ($\chi^2(4) = 9.352$, $p = 0.053$).[62] Again, an alternative alpha-level of 0.06 would not be unreasonable. Therefore, in terms of use of unmodeled knowledge (*C*) as the dependent variable, differences among feedback conditions mattered for slopes and intercepts. The boxplots in Figure 22 illustrate the degree to which intercepts and slopes vary among feedback conditions. In Figure 23 appear fitted growth curves for each of the feedback conditions. The growth curves illustrated in Figure 23 suggest that changes in use of unmodeled knowledge differ at least between feedback conditions 1 and 2 ($\gamma_{group1} = -0.041$ with a 95% confidence interval of [-0.068, -0.014], $\gamma_{group2} = 0.018$ with a 95% confidence interval of [-0.011, 0.047], and $\gamma_{group3} = 0.001$ with a 95% confidence interval of [-0.025, 0.028]). The group that received the least amount of feedback (i.e., group 1, which received no explicit feedback about

[62] If it had been decided that the inclusion of random slopes had not fit the data, then a test of whether random intercepts alone vary by feedback condition would have rejected the null hypothesis. $\chi^2(4) = 10.554$, p = 0.032.

the disordinal interaction) declined in use of unmodeled knowledge over time, while at least one of the groups that received explicit feedback about the disordinal interaction (i.e., group 2) probably increased in its use of that information over time (although not to a statistically significant extent). These findings are consistent with the idea that use of unmodeled knowledge increases with more feedback about the disordinal interaction but decreases with none. Furthermore, the overall growth curve for *C* is negatively sloped (γ = -0.008 with a 95% confidence interval of [-0.026, 0.010]). This finding is inconsistent with overall expectations, but it is not statistically significant.

Criterion-related validity of unmodeled knowledge (r_z). The ICC for this dependent variable equals 0.23, indicating that the person accounts for 23% of the variability in r_z. The software program R, which permits a likelihood ratio comparison of a fully fixed effects model with a random intercepts model, indicates that the random intercepts model is a better fit to the data ($\chi^2(1) = 37.672$, $p < 0.0001$). Therefore, a multilevel model (i.e., at least a random intercepts model) fits the data better than a fully fixed effects regression model does. Using an alpha-level of 0.05, a random slopes (and intercepts) model barely fails to fit the data better than does a random intercepts model based on a log likelihood chi-square test ($\chi^2(2) = 5.804$, p = 0.055). As with the prior analysis of the *C* parameter, an alpha level of 0.06 is not unreasonable here. Therefore, both slopes and intercepts will be said to vary across subjects in terms of their change in use of the criterion related validity of unmodeled knowledge (such knowledge presumably being that of the disordinal interaction). The trellis plot in Figure 24 does suggest variation in intercepts and slopes across subjects. Furthermore, using an alpha level of 0.05, a model that tests whether random slopes and intercepts vary by feedback condition barely fails to improve fit of a random slopes and intercepts model ($\chi^2(4) = 9.404$, $p = 0.052$). Once more, an alpha level of 0.06 would permit one to conclude instead that there are statistically significant slopes and intercepts of growth curves among feedback conditions.[63] Therefore, in terms of use of the criterion-related validity of unmodeled knowledge (r_z) as the dependent variable, differences among feedback conditions mattered for intercepts and slopes. The boxplots in Figure 25 illustrate how intercepts and slopes vary among feedback conditions. In Figure 26 appear fitted growth curves for each of the feedback conditions. The findings for growth trajectories (as well as the aforementioned findings for r_z) mirror those for the *C* parameter, suggesting that the driver for changes in *C* over time was criterion-relevant information. The growth curves illustrated in Figure 26 suggest that slopes differ at least between feedback conditions 1 and 2 (γ_{group1} = -0.044 with a 95% confidence interval of [-0.067, -0.013], γ_{group2} = 0.018 with a 95% confidence interval of [-0.011, 0.046], and γ_{group3} = 0.001 with a 95% confidence interval of [-0.025, 0.027]). The group that received no explicit feedback about the disordinal interaction (group 1) declined in use of criterion-relevant unmodeld knowledge, while a group that did receive such explicit feedback (group 2) increasingly used it. Unexpectedly, the overall growth curve is negative in slope (γ = -0.008 with a 95% confidence interval of [-0.025, 0.010]). However, as said previously for the *C* parameter, the overall slope here is not statistically significant.

Mechanical knowledge (*G*). The ICC for this dependent variable equals 0.26, indicating that the person accounts for 26% of the variability in *G*. The software program R, which permits a likelihood ratio comparison of a fully fixed effects model with a random intercepts model, indicates that the random intercepts model is a better fit to the data ($\chi^2(1) = 45.646$, $p < 0.0001$). Therefore, a multilevel model (i.e., at least a random intercepts model) fits the data better than a fully fixed effects regression model does. However, a random slopes (and intercepts) model does *not* fit the data better than does a random intercepts model based on a log likelihood chi-square test ($\chi^2(2) = 0.534$, $p = 0.766$). Therefore, only intercepts vary across subjects in terms of their

[63] If only random intercepts had fit the data, a test of whether random intercepts vary by feedback condition would have rejected the null hypothesis. $\chi^2(4) = 10.617$, p = 0.031.

change in G (use of knowledge about the mechanical approach). The trellis plot in Figure 27 does suggest variation in intercepts and slopes across subjects, but the variation in slopes is not statistically significant. Furthermore, a model that tests whether random intercepts vary by feedback condition does not improve fit of a random intercepts model ($\chi^2(4) = 1.297$, $p = 0.862$). Therefore, in terms of G, differences among feedback conditions were immaterial. The boxplots in Figure 28 indicates that intercepts, but not slopes, vary among feedback conditions. In Figure 29 appears a fixed effects regression line (i.e. an average across subjects). The slope is negative ($\gamma = -0.048$ but with a 95% confidence interval of [-0.135, 0.039] that contains 0 and positive values), indicating that mechanical knowledge probably worsens on average over time (but not to a statistically significant extent). That finding is consistent with the idea that any machine-like cognitive efficiency of the human mind declines over time with fatigue. Alternatvely, judges with their limited cognitive resources might have opted to increasingly focus more of their attention on unmodeled knowledge and away from the mechanical kind.

Cognitive control (R_s). The ICC for this dependent variable equals 0.27, indicating that the person accounts for 27% of the variability in R_s. The software program R, which permits a likelihood ratio comparison of a fully fixed effects model with a random intercepts model, indicates that the random intercepts model is a better fit to the data ($\chi^2(1) = 43.574$, $p < 0.0001$). Therefore, a multilevel model (i.e., at least a random intercepts model) fits the data better than a fully fixed effects regression model does. Furthermore, a random slopes (and intercepts) model fits the data better than does a random intercepts model based on a log likelihood chi-square test ($\chi^2(2) = 48.924$, $p = 2.378 \times 10^{-11}$). Therefore, both intercepts and slopes vary across subjects in terms of their change cognitive control as measured by R_s. The trellis plot in Figure 30 does suggest variation in intercepts and slopes across subjects. Moreover, a model that tests whether random intercepts and slopes vary by feedback condition improved fit of a random effects model ($\chi^2(4) = 14.687$, $p = 0.005$). Therefore, in terms of R_s as the dependent variable, differences among feedback conditions matter. The boxplots in Figure 31 indicatee that intercepts and slopes do vary among feedback conditions, at least somewhat. In Figure 32 appear growth curves for the feedback conditions (γ_{group1} = 0.207 with a 95% confidence interval of [0.074, 0.340], γ_{group2} = 0.005 with a 95% confidence interval of [-0.136, 0.146], and γ_{group3} = -0.08 with a 95% confidence interval of [-0.209, 0.049]). Groups 1 and 3 differed significantly in growth trajectory. One can see that the feedback group that received the least amount of information (i.e., no explicit information about the existence of the disordinal interaction) was the only feedback group to experience much positive growth in R_s (as well as any statistically significant positive growth). In other words, without information about the disordinal interaction, group 1 became more like a consistently used bootstrapped linear model of their clinical approaches. Since use of the disordinal interaction rather than this alternative prediction strategy is a key to more accurate prediction, group 1's strategy worsened over time. Group 3, which received the most information about the disordinal interaction, experienced negative growth in R_s, which is consistent with the notion that more relevant feedback over time leads to a clinical strategy that is decreasingly like an unvarying bootstrapped linear judgment approach. The overall fixed effect for the growth trajectory was 0.043 with a 95% confidence interval of [-0.038, 0.123]. Contrary to expectations (but not to a statistically significant extent), cognitive control increased over time for subjects on average.

Relative weight (C_{xj}). The ICC for this dependent variable equals 0.35, indicating that the person accounts for 35% of the variability in C_{xj}. The software program R, which permits a likelihood ratio comparison of a fully fixed effects model with a random intercepts model, indicates that the random intercepts model is a better fit to the data ($\chi^2(1) = 82.110$, $p = 0.0001$). Therefore, a multilevel model (i.e., at least a random intercepts model) fits the data better than a fully fixed effects regression model does. Furthermore, a random slopes (and intercepts) model fits the data better than does a random intercepts model based on a log likelihood chi-square test ($\chi^2(2) = 23.691$, $p = 7.172*10^{-6}$). Therefore, both intercepts and slopes vary across subjects in terms of growth in the use of the disordinal interaction as measured by the relative weight for the disordinal interaction. The trellis plot in Figure 33 does suggest variation in intercepts and slopes across subjects. A model that tests whether random intercepts and random slopes vary by feedback condition does *not* improve fit of a random intercepts and slopes model ($\chi^2(4) = 0.927$, $p = 0.921$).[64] Therefore, in terms of the relative weight of the disordinal interaction as the dependent variable, differences among feedback conditions were immaterial. The boxplots in Figure 34 show some variation in intercepts and slopes across feedback conditions, but it is not statistically significant. In Figures 35 and 36 appear, respectively, fitted growth curves (regression lines) for all subjects as well as a single fitted growth curve for subjects overall. The first graph is somewhat ambiguous as to whether the subjects tend to increase or decline in weighing the disordinal interaction over time, although some of the negatively sloped lines are especially steep in comparison to the positively sloped lines generally. The second graph shows the fixed effect growth curve (an average trend ignoring random effects). As one can see, the regression line is negatively sloped ($\gamma = -0.0014$ but with a 95% confidence interval of [-0.009, 0.007] that clearly contains 0 and positive values). Individuals on average declined in use of the disordinal interaction as measured by C_{xj}. However, just under half of the values within the foregoing confidence interval are positive in value, so that declining trend is not statistically significant.

Spring 2010 Results

Judgment criterion-related validity (r_a). The ICC for this dependent variable equals 0, indicating that the person accounts for none of the variability in r_a. The software program R, which permits a likelihood ratio comparison of a fully fixed effects model with a random intercepts model, indicates that the random intercepts model is *not* a better fit to the data (for a comparison of models both with a feedback condition x time interaction, $\chi^2(1) = 3.845*10^{-7}$, $p = 0.9995$; for a comparison of models both without a feedback condition x time interaction, $\chi^2(1) =$

[64] A model that tests whether random intercepts alone (without random slopes) would vary by feedback condition does *not* improve fit of a random intercepts model either ($\chi^2(4) = 0.930$, p = 0.920).

$3.950*10^{-7}$, p = 0.9995). Therefore, a multilevel model (i.e., at least a random intercepts model) does *not* fit the data better than a fully fixed effects regression mode does. Furthermore, a test of a fully fixed effects model with a feedback condition x time interaction barely fails to fit the data better than does a fully fixed effects model without that interaction term ($F(4) = 2.23$, $p = 0.064$). However, one should remember that the choice of using an alpha-level of 0.05 is merely historical convention (Cowles & Davis, 1982), so it may not make sense to reject the idea that feedback conditions make no difference in the change in judgment accuracy over time. As shown in the boxplots of Figure 50, there is some variability in intercepts and slopes across feedback conditions. If one were to ignore the feedback condition x time interaction, then the fixed effect growth curve for everyone would be illustrated as Figure 51 (where slope $b = 0.034$ with a 95% confidence interval of [0.016, 0.052]). This result indicates that on average subjects are learning to predict more accurately over time. If one were to look at growth for each feedback condition separately, then the fitted growth curves would be illustrated as Figure 52 (where slopes b_{group1} = 0.030 with a 95% confidence interval of [-0.001, 0.062], b_{group2} = 0.031 with a 95% confidence interval of [0.001, 0.062], and b_{group3} = 0.040 with a 95% confidence interval of [0.009, 0.071]). As one can see, the main difference among groups would seem to be mainly in intercepts only, and then mainly between group 3 and the other two groups. However, the slope of group 3 is somewhat more steeply positive than the slopes of the other two groups (although it is not statistically signifcant). Also, only group 1 has a confidence interval that contains 0 and negative values, making it more likely that group 1 rather than the other groups actually has a negative slope. These latter two findings are consistent with the idea that less feedback about the disordinal interaction leads to decline in predictive accuracy whereas more feedback about it leads to greater accuracy.

Skill score. The ICC for this dependent variable equals 0.03, indicating that the person accounts for only 3% of the variability in skill score. The software program R, which permits a likelihood ratio comparison of a fully fixed effects model with a random intercepts model, indicates that the random intercepts model is *not* a better fit to the data (for a comparison of models both with a feedback condition x time interaction, $\chi^2(1) = 0.754$, $p = 0.385$; for a comparison of models both without a feedback condition x time interaction, $\chi^2(1) = 0.993$, $p = 0.319$). Therefore, a multilevel model (i.e., at least a random intercepts model) does *not* fit the data better than a fully fixed effects regression model does. Furthermore, a test of a fully fixed effects model with a feedback condition x time interaction fails to fit the data better than does a fully fixed effects model without that interaction term ($F(4) = 1.10$, $p = 0.355$). As shown in the boxplots of Figure 53, there is some variability in intercepts and slopes across feedback conditions, but it is not statistically significant. The single fitted growth curve for everyone would be illustrated as the first plot in Figure 54 (where slope $b = 0.063$ with a 95% confidence interval of [-0.037, 0.114]). The overall conclusion is that for all subjects considered together, there is likely to be a positive change in predictive accuracy as measured by skill score over time.

Unmodeled knowledge (*C*). The ICC for this dependent variable equals 0.18, indicating that the person accounts for 18% of the variability in *C*. The software program R, which permits a likelihood ratio comparison of a fully fixed effects model with a random intercepts model, indicates that the random intercepts model is a better fit to the data ($\chi^2(1) = 46.12$, $p < 0.0001$). Therefore, a multilevel model (i.e., at least a random intercepts model) fits the data better than a fully fixed effects regression model does. Furthermore, a random slopes (and intercepts) model fits the data better than does a random intercepts model based on a log likelihood chi-square test ($\chi^2(2) = 13.761$, $p = 0.001$). Therefore, both intercepts and slopes vary across subjects in terms of their change in use of unmodeled knowledge. The trellis plot in Figure 55 does suggest variation in intercepts and slopes across subjects. However, a model that tests whether random intercepts and random slopes vary by feedback condition does *not* improve fit of a random effects model ($\chi^2(4) = 6.724$, $p = 0.151$ for a random intercepts and slopes model).[65] Therefore, in terms

[65] Even if the best fitting model had been a random intercepts only model, there would not have been a statistically significant finding for feedback conditions ($\chi^2(4) = 7.69$, p = 0.104).

of unmodeled knowledge as the dependent variable, differences among feedback conditions did *not* matter. The boxplots in Figure 56 indicate that intercepts and slopes *do* vary among feedback conditions, although that variation is *not* statistically significant. Figure 57 illustrates that for individual subjects the change in use of unmodeled knowledge over time is predominantly positive. The first graph shows this with individual subjects' regression lines (growth curves). The second graph in Figure 58 shows the positively sloped growth curve for the fixed effect only (i.e. the overall average), where $\gamma = 0.074$ (with a 95% confidence interval of [0.053, 0.094]).

Criterion-related validity of unmodeled knowledge (r_z). The ICC for this dependent variable equals 0.17, indicating that the person accounts for 17% of the variability in r_z. The software program R, which permits a likelihood ratio comparison of a fully fixed effects model with a random intercepts model, indicates that the random intercepts model *is* a better fit to the data ($\chi^2(1) = 44.734$, $p < 0.0001$). Therefore, a multilevel model (i.e., at least a random intercepts model) fits the data better than a fully fixed effects regression model does. Furthermore, a random slopes (and intercepts) model fits the data better than does a random intercepts model based on a log likelihood chi-square test ($\chi^2(2) = 12.588$, $p = 0.0018$). Therefore, both intercepts and slopes vary across subjects in terms of their change in cognitive control. The trellis plot in Figure 59 does suggest variation in intercepts and slopes across subjects. Furthermore, a model that tests whether random intercepts and random slopes vary by feedback condition does *not* improve fit of a random effects model ($\chi^2(4) = 6.12$, $p = 0.1905$ for a random intercepts and slopes model). Therefore, in terms of cognitive control as the dependent variable, differences among feedback conditions are immaterial. The boxplots in Figure 60 indicate that intercepts and slopes do vary among feedback conditions, but that variation is not statistically significant. In Figure 61 appears the overall fixed effect curve fitted to the data (slope $\gamma = 0.066$ with a 95% confidence interval of [0.049, 0.084]). This result indicates that use of criterion-related unmodeled knowledge increased over time for subjects on average (and such increase was statistically singificant). This finding is consistent with the expectation that use of the disordinal interaction would on average increase over time, because all subjects received some information about the disordinal interaction (and most were told explitily that it might have existed or did exist).

Mechanical knowledge (*G*). The ICC for this dependent variable equals 0, indicating that the person accounts for none of the variability in *G*. The software program R, which permits a likelihood ratio comparison of a fully fixed effects model with a random intercepts model, indicates that the random intercepts model is *not* a better fit to the data ($\chi^2(1) = 6.78 * 10^{-7}$, $p = 0.9993$). Therefore, a multilevel model (i.e., at least a random intercepts model) does *not* fit the data better than a fully fixed effects regression model does. Moreover, a model that tested whether regression lines differed by feedback condition did *not* improve fit of a random effects model ($F(4) = 0.618$, $p = 0.650$). The boxplots in Figure 62 indicate that intercepts and slopes do vary among feedback conditions, although that variation is not statistically significant. In Figure 63 appears the fitted fixed growth curve for all subjects showing how mechanical knowledge declines over time (slope $b = -0.079$ with a 95% confidence interval of [-0.141, -0.018]). This statistically significant decline in use of mechanical knowledge is not surprising. One would expect a fatigue effect and a greater focus by most subjects on unmodeled knowledge in lieu of mechanical knowledge in order to maximize accuracy.

Cognitive control (R_s). The ICC for this dependent variable equals 0.27, indicating that the person accounts for 27% of the variability in R_s. The software program R, which permits a likelihood ratio comparison of a fully fixed effects model with a random intercepts model, indicates that the random intercepts model is a better fit to the data ($\chi^2(1) = 99.773$, $p < 0.0001$). Therefore, a multilevel model (i.e., at least a random intercepts model) fits the data better than a fully fixed effects regression model does. Furthermore, a random slopes (and intercepts) model fits the data better than does a random intercepts model based on a log likelihood chi-square test ($\chi^2(2) = 57.977$, $p = 2.573 * 10^{-13}$). Therefore, both intercepts and slopes vary across subjects in terms of their change in cognitive control. The trellis plot in Figure 64 does suggest variation in intercepts and slopes across subjects. Furthermore, a model that tests whether random intercepts and random slopes vary by feedback condition improves fit of a random effects model ($\chi^2(4) = 11.608$, $p = 0.021$ for a random intercepts and slopes model). Therefore, in terms of cognitive control as the dependent variable, differences among feedback conditions matter. The boxplots in Figure 65 indicate that intercepts and slopes do vary among feedback conditions. In Figure 66 appear fitted growth curves for R_s. As one can see, there are declines over time in R_s for each feedback condition ($\gamma_{group1} = -0.151$ with a 95% confidence interval of [-0.229, -0.073], $\gamma_{group2} = -0.108$ with a 95% confidence interval of [-0.184, -0.031], and $\gamma_{group3} = -0.230$ with a 95% confidence interval of [-0.307, -0.153]). The overall fixed effects growth curve (slope $\gamma = -0.162$ with a 95% confidence interval of [-0.202, -0.123]). A decline in R_s is related to greater use of unmodeled knowledge. Past research suggests that there would be greater use of unmodeled knowledge (particularly the disordinal interaction) over time as people receive more feedback about that knowledge. It is expected that the group that was explicitly told in words that the disordinal interaction existed and which also received population-level graphical feedback about the interaction (feedback group 3) saw the steepest decline in R_s.

Relative weight (*C_{xj}*). The ICC for this dependent variable equals 0.26, indicating that the person accounts for 26% of the variability in C_{xj}. The software program R, which permits a likelihood ratio comparison of a fully fixed effects model with a random intercepts model, indicates that the random intercepts model is a better fit to the data ($\chi^2(1) = 81.981$, $p < 0.0001$). Therefore, a multilevel model (i.e., at least a random intercepts model) fits the data better than a fully fixed effects regression model does. Furthermore, a random slopes (and intercepts) model fits the data better than does a random intercepts model based on a log likelihood chi-square test ($\chi^2(2) = 74.458$, $p = 2.2*10^{-16}$). Therefore, both intercepts and slopes vary across subjects in terms of their change in weighting of the disordinal interaction. The trellis plot in Figure 67 does suggest variation in intercepts and slopes across subjects. Furthermore, a model that tests whether random intercepts and random slopes vary by feedback condition improves fit of a random effects model ($\chi^2(4) = 12.473$, $p = 0.014$ for a random intercepts and slopes model). Therefore, in terms of relative weight as the dependent variable, differences among feedback conditions matter. The boxplots in Figure 68 indicate that intercepts and slopes do vary among feedback conditions. In Figure 69 appear fitted growth curves for C_{xj}. As one can see, the relative weight for each feedback condition increases over time ($\gamma_{group1} = 0.013$ with a 95% confidence interval of [0.0001, 0.026], $\gamma_{group2} = 0.018$ with a 95% confidence interval of [0.005, 0.031], and $\gamma_{group3} = 0.026$ with a 95% confidence interval of [0.013, 0.039]). The overall fitted fixed effect slope is positive 0.019 with a 95% confidence interval of [0.012, 0.026]. These results (all positive and statistically significant) are not surprising if one expects greater use of the disordinal interaction over time for all groups. The group that was explicitly told in words that the disordinal interaction existed and which also received population-level graphical feedback about the interaction (feedback group 3) saw the steepest increase in the relative weight, followed by the group that was merely told that the disordinal interaction might exist (group 2). Group 1, which saw the shallowest increase, received no explicit wording about the disordinal interaction (but did receive graphical feedback about it). These results are expected if more feedback about the interaction causes subjects to use the interaction more often in making predictions.

Discussion

On the whole, Fall 2009 and Spring 2010 results are consistent with the idea that, when given sufficient feedback about the disordinal interaction, subjects learned to better leverage unmodeled knowledge over time to improve the accuracy of their judgments. It is true that r_a declines over time when subjects in the Fall 2009 design are looked at as a whole, but that trend is statistically insignificant. Also, model testing indicates that subjects should be considered separately as individuals (for both slope and intercept). The change in r_a over time for the Spring 2010 design was positive whether one did or did not consider the feedback groups separately. When one does consider them separately, the only group that received explicit wording that the disordinal interaction existed (i.e., group 3) experienced somewhat better positive growth than the other 2 feedback groups. When the measure of accuracy is skill score, evidence even more strongly shows improvement in clinical accuracy over time. For the Fall 2009 experiment, slopes were positive for only those groups who received explicit information about the disordinal interaction (i.e. groups 2 and 3). The fitted slope for skill score growth was positive for everyone in Spring 2010.

The Lens Model parameters underlying holistic accuracy as measured by r_a are generally supportive of the hypothesis that people will learn to increasingly rely upon the disordinal interaction as they continue to receive task information feedback. Use of unmodeled knowledge (measured by C) in Fall 2009 increased for only the 2 groups that received explicit information about the disordinal interaction (i.e., groups 2 and 3). For Spring 2010, usually the C parameter increased for people over time. The encouraging findings for changes in the C parameter for Fall 2009 replicated for r_z. In other words, not only did use of unmodeled knowledge increase over time for only those groups receving explicit feedback about the disordinal interaction, but so too did use of that part of unmodeled knowledge which is predictive of the job performance. For Spring 2010, the criterion relevant validity of C (i.e., r_z) increased over time for all subjects.

Changes in R_s due to training indicate that people were looking for new information outside of a purely linear model in order to try to better predict the dependent variable. Recall that C is the correlation between the residuals of the environment (the purely mechanical equation) and the residuals of the judge's bootstrapped model. It is $\sqrt{(1-R_s{}^2)}$, essentially the opposite of R_s, which is that part of C that is specific to the judge's bootstrapped model. Therefore, decreases in R_s contribute positively to C. For Fall 2009, the group that received no explicit feedback about the disordinal interaction (group 1) saw sharp increases in R_s. Group 3, which received the most information explicitly about the disordinal interaction, saw a fairly rapid decline in R_s across time. While group 2 (which did receive explicit information about the disordinal interaction) saw an increase in R_s over time, that increase was relatively shallow. For Spring 2010, all feedback conditions saw decreases in R_s, but the steepest decreases occurred for group 3, the only group to receive explicit wording that the disordinal interaction actually existed.

Changes in the *G* parameter in Fall 2009 and Spring 2010 indicate that use of mechanical knowledge declined over time for all subjects. This decrease might explain why r_a declined for all subjects in Fall 2009. The decline in *G* may indicate that subjects were becoming too fatigued to maintain a consistent prediction strategy akin to a mechanical process. It also is possible that over time subjects were increasingly diverting their focus away from a linear mechanical approach and toward a configural judgment strategy.

Results for the C_{xj} relative weight for the disordinal interaction were mixed. The fixed effect for all subjects in Fall 2009 indicates that the weight given to the disordinal interaction declined over time. However, sampling error might explain this result. In contrast, for Spring 2010, the growth curve for C_{xj} was positively sloped and statistically significant for all groups and especially for group 3 (the only group to receive explicit wording that the disordinal interaction existed). There is some evidence that with training, people will increasingly rely on the disordinal interaction to maximize judgment accuracy.

Change in Confidence

The third research question concerns the extent (if any) to which assessors can learn to become less overconfident in their judgment strategies. There are 2 confidence dimensions. Confidence can be absolute or relative. Confidence can be overall or measured after every block. That is, confidence can be absolute and overall, absolute and measured after every block, relative and overall, or relative and measured after every block. Absolute confidence refers to self-reported confidence without regard to how the subject thought that any other participant performed in making predictions of job performance. Relative confidence does specifically consider others' performances. Overall confidence refers to self-reported confidence for the entire task (all blocks considered together). Alternatively, confidence can be measured after every block.

Based on previous research, one generally would expect all types of confidence to decline for those who receive feedback about the nature of the task (especially the disordinal interaction). One would expect those who receive less feedback to experience less of a decline or perhaps an increase in confidence. If people generally do not think about how accurately they engage in prediction of human behavior, then as people increasingly realize the complexity of the task involved, people are expected to question their ability to perform well. When people do not receive as much feedback, the complexity of the task becomes less apparent. Therefore, people may lack any counterweight to entrenched and often erroneous beliefs about their acumen in both a relative and absolute sense.

In Figure 37 appears one line graph depicting group changes in absolute confidence for the Fall 2009 study overall as well as after every block. Figure 38 depicts group changes in relative confidence for the study overall as well as after every block. The markers that are not connected by lines represent self-reported confidence for the task overall before any predictions were made and after all predictions were made. (See the Academic Experience and Information Form in Appendix L for the questions about confidence prior to making any predictions and Appendix Q for the questions about confidence after making all predictions). The markers that are connected by lines represent self-reported confidence measured after every block. (See Appendix P for the questions about confidence after every block). The graphs of absolute and relative confidence in the Spring 2010 study (see Figures 70 and 71) would be interpreted in a similar way.

Statistical results used to analyze overall confidence are summarized in Table 8. Statistical results used to analyze confidence measured after every block appear in Table 9. HLM was used to conduct these analyses of confidence measured after every block and is the basis for discussion of them. Reasons for employing HLM as well as the nature of the HLM procedures mirror those for employing HLM to measure change in clinical validity (see "Change in Clinical Accuracy", above).

Fall 2009 Results

Absolute confidence.

Overall confidence. Figure 37 indicates that overall confidence decreased across all 3 feedback conditions. However, these results are not all statistically significant (where $\alpha = 0.05$). Where negative *d*-values indicate a decrease in confidence, for feedback group 1 $d = -0.25$ with a 95% confidence interval of [-0.72, 0.23] ($N = 30$). For feedback group 2, $d = -0.82$ with a 95% confidence interval of [-1.27, -0.37] ($N = 34$). For feedback group 3, $d = -0.61$ with a 95% confidence interval of [-0.87, -0.35] ($N = 38$). Since only groups 2 and 3 had confidence intervals that did not contain 0, one would conclude that there was a statistically significant decrease in confidence for only those feedback groups that received explicit information about the disordinal interaction. For group 1, $t(29) = 1.07$, $p = 0.293$ (with a mean of differences equal to -0.2333). For group 2, $t(33) = 3.69$, $p = 0.0008$ (with a mean of differences equal to -0.7059). For group 3, $t(37) = 4.714$, $p = 3.401*10^{-5}$ (with a mean of differences equal to -0.6974). Regarding the overall confidence for the population, significance tests indicate that absolute confidence decreased for those that received explicit feedback about the disordinal interaction (groups 2 and 3) but not for the group that did not receive it (group 1). This is fully consistent with the conclusions from looking at confidence intervals of *d* values.

Confidence measured after every block. The ICC for absolute confidence equals 0.64, indicating that the person accounts for 64% of the variability in absolute confidence. The software program R, which permits a likelihood ratio comparison of a fully fixed effects model with a random intercepts model, indicates that the random intercepts model is a better fit to the data ($\chi^2(1) = 270.772$, $p < 0.0001$). Therefore, a multilevel model (i.e., at least a random intercepts model) fits the data better than a fully fixed effects regression model does. Furthermore, a random slopes (and intercepts) model fits the data better than does a random intercepts model based on a log likelihood chi-square test ($\chi^2(2) = 23.308$, $p = 8.685*10^{-6}$). Therefore, both intercepts and slopes vary across subjects in terms of their change in absolute confidence. The trellis plot in Figure 39 does suggest variation in intercepts and slopes across subjects. Furthermore, a model that tests whether random intercepts and random slopes vary by feedback condition improves fit of a random effects model ($\chi^2(4) = 11.89$, $p = 0.018$ for a random intercepts and slopes model). Therefore, in terms of absolute confidence as the dependent variable, differences among feedback conditions mattered. The boxplots in Figure 40 indicate that intercepts and slopes do vary among feedback conditions. In particular, feedback group 1 (which received no explicit information about the disordinal interaction) was quite discrepant in intercept and slope from the other two feedback conditions which did receive that information (although groups 1 and 3 possess overlapping confidence intervals). Its intercept was smaller, and its slope was greater. In Figure 41 appear fitted growth curves for each of the feedback conditions (with slopes $\gamma_{group1} = 0.090$ with a 95% confidence interval of [-0.010, 0.190], $\gamma_{group2} = -0.154$ with a 95% confidence interval of [-0.260, -0.048], and $\gamma_{group3} = -0.074$ with a 95% confidence interval of [-0.171, 0.024]). The fitted growth curve for feedback group 1 was steeply positive in slope in comparison to the negative slopes for the other fitted growth curves. It appears that lack of feedback about the disordinal interaction most likely led to increased confidence over time (in group 1), whereas provision of that feedback information most likely led to decreased confidence over time (in groups 2 and 3).

Relative confidence.

Overall confidence. Figure 38 indicates that overall confidence decreased across all 3 feedback conditions. However, not all of these results are statistically significant (where $\alpha = 0.05$). Where negative d-values indicate a decrease in confidence, for feedback group 1 $d = -0.40$ with a 95% confidence interval of [-0.81, 0.02] ($N = 28$). For feedback group 2, $d = -0.48$ with a 95% confidence interval of [-0.89, -0.06] ($N = 35$). For feedback group 3, $d = -0.48$ with a 95% confidence interval of [-0.75, -0.22] ($N = 37$). Since only groups 2 and 3 had confidence intervals that did not contain 0, one would conclude that there was a decrease in confidence for only those feedback groups that received explicit information about the disordinal interaction. For group 1, $t(27) = 1.97$, $p = 0.06$ (with a mean of differences equal to -0.3214). Once again, one should remember that the choice of using an alpha-level of 0.05 is merely historical convention (Cowles & Davis, 1982). For group 2, $t(34) = 2.32$, $p = 0.03$ (with a mean of differences equal to -0.4286). For group 3, $t(36) = 3.72$, $p = 0.0007$ (with a mean of differences equal to -0.5676). Regarding the overall confidence for the population, significance tests indicate that relative confidence decreased only for the groups that received explicit feedback about the disordinal interaction (groups 2 and 3). This is fully consistent with the conclusions from looking at confidence intervals for *d*-values.

Confidence measured after every block. The ICC for relative confidence equals 0.71, indicating that the person accounts for 71% of the variability in relative confidence. The software program R, which permits a likelihood ratio comparison of a fully fixed effects model with a random intercepts model, indicates that the random intercepts model is a better fit to the data ($\chi^2(1) = 339.884$, $p < 0.0001$). Therefore, a multilevel model (i.e., at least a random intercepts model) fits the data better than a fully fixed effects regression model does. Furthermore, a random slopes (and intercepts) model fits the data better than does a random intercepts model based on a log likelihood chi-square test ($\chi^2(2) = 33.51$, $p = 5.29*10^{-8}$). Therefore, both intercepts and slopes vary across subjects in terms of their change in relative confidence. The trellis plot in Figure 42 does suggest variation in intercepts and slopes across subjects. However, a model that tests whether random intercepts and random slopes vary by feedback condition does not improve fit of a random effects model ($\chi^2(4) = 8.372$, $p = 0.079$ for a random intercepts and slopes model). Therefore, in terms of relative confidence as the dependent variable, one would conclude that differences among feedback conditions did not matter based on an alpha level of 0.05. On the other hand, a model that tests whether random intercepts alone (without random slopes) would vary by feedback condition does improve fit of a random intercepts model ($\chi^2(4) = 14.169$, $p = 0.0068$). Once again, one should remember that the choice of using an alpha level of 0.05 is merely historical convention (Cowles & Davis, 1982). Therefore, one should look at the failure to reject the null hypothesis for the test of whether random intercepts and slopes vary by feedback condition (where $p = 0.079$) with caution. The boxplots in Figure 43 indicate that intercepts and slopes do vary among feedback conditions. In particular, feedback group 1 (which received no explicit information about the disordinal interaction) was discrepant in intercept and slope from the other two feedback conditions (which did receive that information). Its intercept was lower, and its slope was arguably greater given its positive skew and the negative skew for group 3. In Figures 44 and 45, respectively, appear a fitted fixed effect growth curve for subjects in general and fitted growth curves for each of the feedback conditions. The slope of the fixed effect line is negative ($\gamma = -0.052$ with a 95% confidence interval of [-0.108, -0.009]). On average, subjects declined in relative confidence, and this decline was statistically significant. The fitted growth curve for feedback group 1 was steeply positive in slope in comparison to the negative slopes for the other fitted growth curves ($\gamma_{group1} = 0.05$ with a 95% confidence interval of [-0.054, 0.112], $\gamma_{group2} = -0.110$ with a 95% confidence interval of [-0.201, -0.025], and $\gamma_{group3} = -0.104$ with a 95% confidence interval of [-0.179, -0.014]), but the slope for group 1 was not statistically significant. Provision of feedback about the disordinal interaction information led to decreased confidence over time (in groups 2 and 3), and that decrease was statistically significant.

Spring 2010 Results

Absolute confidence.

Overall confidence. Figure 70 indicates that overall confidence decreased across all 3 feedback conditions, and these results are statistically significant (where $\alpha = 0.05$). Where negative d-values indicate a decrease in confidence, for feedback group 1 $d = -0.76$ with a 95% confidence interval of [-1.26, -0.26] ($N = 33$). For feedback group 2, $d = -0.58$ with a 95% confidence interval of [-0.965, -0.196] ($N = 41$). For feedback group 3, $d = -0.60$ with a 95%

confidence interval of [-0.99, -0.21] ($N = 40$). Since all groups had confidence intervals that did not contain 0, one would conclude that there was a decrease in confidence for all feedback groups. For group 1, $t(32) = 3.12$, $p = 0.004$ (with a mean of differences equal to -0.606). For group 2, $t(40) = 3.05$, $p = 0.004$ (with a mean of differences equal to -0.512). For group 3, $t(39) = 3.10$, $p = 0.004$ (with a mean of differences equal to -0.4625). Regarding the overall confidence for the population, significance tests indicate that absolute confidence decreased for all groups. This is fully consistent with the conclusions from looking at confidence intervals of the *d*-values.

Confidence measured after every block. The ICC for absolute confidence equals 0.57, indicating that the person accounts for 57% of the variability in absolute confidence. The software program R, which permits a likelihood ratio comparison of a fully fixed effects model with a random intercepts model, indicates that the random intercepts model is a better fit to the data ($\chi^2(1) = 284.162$, $p < 0.0001$). Therefore, a multilevel model (i.e., at least a random intercepts model) fits the data better than a fully fixed effects regression model does. Furthermore, a random slopes (and intercepts) model fits the data better than does a random intercepts model based on a log likelihood chi-square test ($\chi^2(2) = 26.596$, $p = 1.678*10^{-6}$). Therefore, both intercepts and slopes vary across subjects in terms of their change in absolute confidence. The trellis plot in Figure 72 does suggest variation in intercepts and slopes across subjects. A model that tests whether random intercepts and random slopes vary by feedback condition does *not* improve fit of a random effects model ($\chi^2(4) = 8.724$, $p = 0.068$ for a random intercepts and slopes model). Nevertheless, one should be willing to look at feedback conditions separately given the low p-value and the fact that 0.05 alpha levels on which many significance tests are based are due solely to historical convention (Cowles & Davis, 1982). Based on the orthodox view of 0.05 alpha levels, one would conclude that differences among feedback conditions did *not* matter. The boxplots in Figure 73 indicate that intercepts and slopes do vary among feedback conditions, but that variation is not statistically significant based on a 0.05 alpha level. Figure 74 shows that a fixed effects regression line has a positive slope ($\gamma = 0.017$ with a 95% confidence interval of [-0.025, 0.060]), indicating that on average absolute confidence increases over time (but not in a statistically significant way). If one were to look at groups separately, one would find that group 1 maintained the highest level of confidence during the study period (see Figure 75), although group 3's average confidence level approaches it over time (slopes $\gamma_{group1} = 0.018$ with a 95% confidence interval of [-0.054, 0.096], $\gamma_{group2} = -0.03$ with a 95% confidence interval of [-0.101, 0.044], and $\gamma_{group3} = 0.063$ with a 95% confidence interval of [-0.011, 0.134]; see Figure 75). Group 1's confidence level declined over time, but that decline was not statistically significant. It was unexpected that group 3 (which receives explicit wording that the disordinal interaction exists) would increase in confidence over time, but such a finding was not statistically significant. As Figure 75 illustrates, group 1 (which receives the least amount of feedback about the disordinal interaction) remained the most confident group. Group 2 (which is told that the disordinal interaction may exist) declined in confidence. Overall, the findings were somewhat consistent with the notion that lack of feedback either contributes to or at least fails to reduce overconfidence and that more feedback reduces it.

Relative confidence.

Overall confidence. Figure 71 indicates that overall confidence decreased across all 3 feedback conditions. However, not all of these results are statistically significant (where $\alpha = 0.05$). Where negative d-values indicate a decrease in confidence, for feedback group 1 $d = -0.43$ with a 95% confidence interval of [-0.92, 0.07] ($N = 32$). For feedback group 2, $d = -0.42$ with a 95% confidence interval of [-0.71, -0.13] ($N = 41$). For feedback group 3, $d = -0.50$ with a 95% confidence interval of [-0.75, -0.25] ($N = 41$). Since only groups 2 and 3 had confidence intervals that did not contain 0, one would conclude that there was a decrease in confidence for only those feedback groups that were explicitly told that the disordinal interaction might or does exist. For group 1, $t(31) = 1.75$, $p = 0.09$ (with a mean of differences equal to -0.375). For group 2, $t(40) = 2.97$, $p = 0.005$ (with a mean of differences equal to -0.415). For group 3, $t(40) = 4.05$, $p =$

0.0002 (with a mean of differences equal to -0.512). Regarding the overall confidence for the population, significance tests indicate that relative confidence decreased only for the groups that were told that the interaction might or did exist (groups 2 and 3). This is fully consistent with the conclusions from looking at confidence intervals for *d*-values.

Confidence measured after every block. The ICC for relative confidence equals 0.66, indicating that the person accounts for 66% of the variability in relative confidence. The software program R, which permits a likelihood ratio comparison of a fully fixed effects model with a random intercepts model, indicates that the random intercepts model is a better fit to the data ($\chi^2(1) = 414.05$, $p < 0.0001$). Therefore, a multilevel model (i.e., at least a random intercepts model) fits the data better than a fully fixed effects regression model does. Furthermore, a random slopes (and intercepts) model fits the data better than does a random intercepts model based on a log likelihood chi-square test ($\chi^2(2) = 19.808$, $p = 4.998 * 10^{-5}$). Therefore, both intercepts and slopes vary across subjects in terms of their change in relative confidence. The trellis plot in Figure 76 does suggest variation in intercepts and slopes across subjects. However, a model that tests whether random intercepts and random slopes vary by feedback condition does *not* improve fit of a random effects model ($\chi^2(4) = 7.20$, $p = 0.126$ for a random intercepts and slopes model).[66] Therefore, in terms of relative confidence as the dependent variable, one would conclude that differences among feedback conditions did not matter. The second through fourth trellis plots in Figure 76 do not seem to indicate any clear pattern of intercepts and slopes across the feedback conditions. The boxplots in Figure 77 indicate that intercepts and slopes do vary among feedback conditions, but this variation was not statistically significant. In Figure 78 appears the negatively sloped fitted growth curve for the fixed effect (i.e. the average for all subjects; $\gamma = -0.01$ with a 95% confidence interval of [-0.044, 0.026]). Thus, relative confidence on average declined (although such finding might not have been statistically significant).

[66] If the best-fitting model had been just a random intercepts one, then it would not have been further improved by adding the interaction involving feedback condition ($\chi^2(4) = 7.87$, p = 0.097).

Discussion

For confidence levels measured after every block, many group confidence intervals contained 0 and/or overlapped with point estimates and intervals for other groups. This makes it difficult to talk meaningfully about the comparisons of differences in growth trajectories for confidence. Nevertheless, and especially those from Fall 2009, the results generally support the idea that people's confidence in their holistic judgment approaches declines with the receipt over time of criterion-relevant task information feedback (in this case, information about the disordinal interaction) and increases over time when that feedback is absent. This is true for absolute, relative, and overall confidence as well as confidence measured after every block. In Fall 2009, both overall absolute and overall relative confidence experienced statistically significant declines for the only 2 feedback groups that received explicit information about the disordinal interaction (i.e., groups 1 and 2). For the same dataset, both absolute and relative confidence measured after every block declined for groups 2 and 3 but increased for group 1. (This assumes an alpha level of about 0.08 for relative confidence measured after every block.) In Spring 2010, overall absolute and overall relative confidence declined for all groups, except that the decline for the overall relative confidence of group 1 (the only group which received no explicit wording about the disordinal interaction) was not statistically significant. For absolute confidence measured after every block in Spring 2010, group 3 unexpectedly increased in confidence. Nevertheless, group 1 expectedly increased in confidence while group 2 declined in it. Since the feedback x time interaction was not close to being statistically significant based on an alpha level of 0.05 for relative confidence measured after every block in Spring 2010, no groups slopes were depicted or reported. However, it should be noted that had they been reported, the group 1 slope would have been positive and the group 2 and 3 slopes negative. In summary, the studies taken as a whole generally support the ideas that judges can be debiased (be made less confident) as more criterion-relevant task information feedback is provided and that judges will either become more biased or less unbiased without that information.

Relationship of Change in Clinical Accuracy to Change in Confidence

The fourth research question concerns the relationship (if any) of any changes in predictive accuracy to any changes in overconfidence.

Fall 2009 Results

As indicated earlier, there is a positive relationship of skill scores to time when explicit feedback about the disordinal interaction is received (group 3 and possibly group 2). The relationship is negative (and statistically significant) when no such explicit feedback is received (group 1; see Table 7). When r_a is the measure of performance accuracy, there is a general decline in accuracy over time, although it is not statistically significant (see Table 7). When confidence is measured overall (before any predictions are made and after all predictions are made), the evidence indicates for both absolute and relative confidence that absolute and relative confidence are negatively related to time for all feedback groups (see Table 8). However, those findings are statistically significant for only those groups explicity told about the disordinal interaction (groups 2 and 3). When absolute or relative confidence is measured after every block, then there is a negative relationship of both absolute and relative confidence to time when people receive explicit feedback about the disordinal interaction (groups 2 and 3; see Table 9). The relationship is positive when no such explicit feedback is received (group 1; see Table 9). Although the evidence is not perfectly clear, some of it does indicate that when one measures judgment validity in terms of not just rank order or shape similarity (r_a) but also in terms of elevation and scatter (skill score), then improved judgment accuracy is associated with declining confidence in both an absolute and relative sense. Furthermore, there is some indication that those who receive more feedback information about the disordinal interaction improve their accuracy (particularly as measured by skill score) while becoming less confident. Those who receive less information about the disordinal interaction decline in accuracy (again, particularly as measured by skill score) while possibly becoming more confident (although significance levels make the determination of change in confidence at best speculative).

Spring 2010 Results

As discussed previously, there is a positive relationship of r_a and skill score to time. For skill score, the positive relationship is statistically nonsignificant for subjects on the whole (see Table 18). For r_a, this finding is statistically significant only for those groups who received explicit wording about the possible or actual existence of the disordinal interaction (groups 2 and 3).   Confidence generally declined over time. It experienced statistically significant decline for overall absolute confidence for all three groups and statistically significant decline for overall relative confidence for the only two groups that received explicit wording about the possible or actual existence of the disordinal interaction (groups 2 and 3). For group 1, the decline was not statistically significant. With respect to confidence measured after every block, relative confidence declined for subjects on the whole. Nevertheless, this decline was not statistically significant. Relative confidence measured after every block increased for groups 1 and 2 but decreased for group 2. However, none of these trends was statistically significant. To the extent that there were statistically significant results, judgment accuracy (particularly as measured by r_a) increased over time while confidence (particularly overall confidence, both absolute and relative) declined over time.

Individual Differences

The fifth research question concerns individual differences (if any) that define those individuals who predict, and learn to predict, most accurately. One way to address this issue is to employ hierarchical regression modeling. The models involved may be multilevel or fully fixed effects OLS ones. In the typical multilevel model, an individual difference variable would be added as a first-level variable, because individuals are what are being are nested within a second-level grouping variable. In a multilevel model for longitudinal data, an individual difference variable is added as second-level (i.e., between-subjects) variable, because individuals *are* the second-level grouping variable within which time points are nested. Tables 11 and 12 (for Fall 2009) and Tables 20 and 21 (for Spring 2010) list all variables entered as a main effect in their own exclusive version of the model that best fits the data when either r_a or skill score is the dependent variable measuring predictive accuracy. These tables indicate which individual difference variables had statistically significant coefficients ("!") and which individual differences led to a model that fit the data better than a model without those individual differences based on a log likelihood chi-square test ("*" or "**"). That is, each individual difference was assessed for whether it had a statistically significant, independent effect on clinical accuracy and whether that effect significantly added, beyond time, to prediction of clinical accuracy. When multi-level models are involved, the probability values for these assessments may differ (see Tables 11 and 12). When models are general OLS ones, then those determinations are redundant, because they will share the same probability value (see Tables 20 and 21). Presumably, this difference between OLS and multilevel models is related to the way in which multilevel modeling partitions variability between levels and thereby affects standard errors.

Furthermore, one can look at between-subjects correlations to assess the importance of individual differences to accuracy. While the analysis using regression models looks at incrementing prediction beyond change over time, this correlational analysis looks at differences among people merely with static "snapshots." In other words, the first set of analyses ask, "Does this individual difference predict change in accuracy beyond the passage of time?", and the second set of analyses ask, "Does this individual difference help to identify those people who will predict more accurately?" In Table 10 (for Fall 2009) and Table 19 (for Spring 2010) appear correlations among all of the variables analyzed in the experiments, only some of which are meaningfully considered individual differences (i.e. cognitive ability, gender, personality, interest, and experiential measures). The only category of individual difference that did not have a statistically significant relationship to a measure of predictive accuracy (based on criteria explained below) was gender. This was unexpected given that the feedback was very spatial in nature and previous empirical findings that on average males outperform females on highly spatial tasks.

Fall 2009 Results

Judgment criterion-related validity (r_a). The following individual difference variables had statistically significant coefficients: ACT composite score, ACT English score, ACT English Writing score, ACT Reading score, ACT Science Reasoning score, cumulative college GPA, high school rank, extroversion/positive emotionality (as measured by the Goldberg and IPC-7 inventories, respectively), conventional interests (as measured by the RIASEC Interest Profiler), and the aggregate length of prior training to make predictions of human behavior (measured in weeks). Of those variables, and based on log likelihood chi-square tests indicating whether a model with the individual difference included would result in better model fit, extroversion/positive emotionality, conventional interests, and aggregate length of prior training to make predictions of human behavior (measured in weeks) added significantly to the prediction of within-person growth in clinical criterion-related validity over time (see Table 11). The coefficients for extroversion and positive emotionality were both negative, indicating that being introverted contributed to more accurate prediction. Conventional interests were positively related to more accurate prediction. Not surprisingly, those who like accuracy, efficiency, and working with numbers perform better. The aggregate length of prior training to make predictions of human behavior (measured in weeks) was negatively related to accuracy. If previous training generally instilled poor judgment strategies (e.g. use of faulty heuristics and biases), then it might have a detrimental impact on performance in the experiments.

Based on between-subjects correlational relationships, there are several interesting findings. All sections of the ACT (verbal, mathematical, writing, reading comprehension, and scientific reasoning ability) as well as ACT composite score were significantly positively related to greater accuracy ($p < 0.05$) as measured by a measure of r_a that possibly reflects the receipt of experimental training (i.e., all measures of r_a except $r_a 1$; see Table 10). Statistically significant correlations ranged from 0.21 to 0.39 ($N = 111$; see Table 10). To a lesser extent, there were similar findings for cumulative GPA ($r = 0.18$, 0.23, and 0.24; $p < 0.05$; $N = 134$) and to an even lesser extent high school rank ($r = 0.20$ and 0.30; $p < 0.05$; $N = 97$; see Table 10). Total academic hours did not bear a significant relationship to any r_a metrics (see Table 10). It is reasonable to conclude that higher cognitive ability leads to better prediction of rank order.

Several personality dimensions had statistically significant relationships with r_a. At some point during or just after training, extroversion ($r = -0.19$ and -0.23; $p < 0.05$; $N = 142$), positive emotionality ($r = -0.24$; $p < 0.05$; $N = 142$), conscientiousness ($r = -0.17$, -0.20; $p < 0.05$; $N = 142$), and conventionality (lack of openness; $r = -0.17$; $p < 0.05$; $N = 142$) had a negative relationship to r_a (see Table 10). Intellectance ($r = 0.18$; $p < 0.05$; $N = 142$) and agreeableness ($r = 0.23$; $p < 0.05$; $N = 142$) had positive relationship to an r_a metric, but the relationship for agreeableness was with r_a prior to subjects' receipt of any feedback (i.e. was with $r_a 1$; see Table 10). This evidence suggests that those who are more introverted, less conscientious, and more open to new experience predict rank order better with training. Without training, agreeable people would be expected to predict more accurately. Except for conscientiousness, these findings are consistent with the notion that learning to better predict rank order requires careful consideration of feedback. The between-subjects finding for conscientiousness is interesting and unexpected. It might simply be due to sampling or other error. Alternatively, more conscientious people might tend to be so thorough that they do not simply follow suggestions in feedback to utilize the disordinal interaction. As discussed earlier, empirical evidence shows that even when researchers explicitly tell subjects that perfect prediction is impossible, and that even when researchers generally inform subjects about the probabilistic nature of a judgment task, subjects' judgment behavior does not improve (see Brehmer & Kuylenstierna, 1978, 1980; Johnansson & Brehmer, 1979). As also discussed earlier, people stubbornly view the use of compromise strategies like unit weighting as an unacceptable introduction of error (Einhorn, 1986). The data displayed as feedback in the Fall 2009 experiment, including the graphs of the disordinal interaction, did suffer from some degree of noise even though that noise was somewhat deliberately constrained (see Appendix E). Perhaps more conscientious people under the experimental learning conditions, in an attempt to achieve greater accuracy, over-thought the judgment task and refused to accept the suggestion of using a disordinal interaction.

Interests also had statistically significant relationships to a measure of r_a that potentially reflects training feedback. Investigative interests ($r = 0.18$; $p < 0.05$; $N = 142$) and conventional interests ($r = 0.20$; $p < 0.05$; $N = 142$) were positively related to an r_a metric, while enterprising interests ($r = -0.18, -0.23$; $p < 0.05$; $N = 142$) were negatively related to an r_a metric (see Table 10). In connection with training, more investigative and conventional and less enterprising people better predict rank order (see Table 10). It is not surprising that people who like to solve problems, who are efficiency-oriented and who are less focused on influencing others will be more accurate in terms of r_a.

In terms of prior experience making predictions of human behavior, only length of previous training ($r = -0.19$; $p < 0.05$; $N = 142$) and amount of prior formal education in judgment and decision making ($r = -0.20$; $p < 0.05$; $N = 142$) bore a significant relationship to r_a at some point during training (see Table 10). The relationships were unexpectedly negative. If the results are not merely due to sampling or other error, then perhaps prior training, experience, and education in judgment were of poor quality and actually instilled tendencies toward suboptimal judgment strategies.

Skill score. The following individual difference variables had statistically significant coefficients: ACT composite score, ACT English score, ACT English Writing score, ACT Reading score, ACT Science Reasoning score, conventionality (lack of openness as measured by the IPC7 inventory), realistic interests, investigative interests, and number of outcomes for which predictions were made in the past in a formal way. Based on log likelihood chi-square tests indicating whether a model with the individual difference included would result in better model fit, all of those variables add significantly to the prediction of within-person growth in clinical accuracy over time (see Table 12). The coefficients for all of the ACT sections, realistic interests, investigative interests, and number of outcomes for which predictions were made in the past in a formal way all were positive in sign. The coefficient for conventionality was negative in sign. It is not surprising that people higher in cognitive ability, who are practical, who like to explore, who have more experience making predictions in a formal way, and who are more open minded to new ideas would make more accurate predictions.

Based on between-subjects correlational relationships, there are several interesting findings. All sections of the ACT (verbal, mathematical, writing, reading comprehension, and scientific reasoning ability) as well as ACT composite score were significantly positively related to greater accuracy ($p < 0.05$) as measured by a skill score metric that possibly reflects experimental training (i.e., all measures of skill score except *SS1*; see Table 10). Statistically significant correlations ranged from 0.22 to 0.30 ($N = 111$). To a lesser extent, there were similar findings for cumulative GPA ($r = 0.18$; $p < 0.05$; $N = 134$; see Table 10). Neither high school rank nor total academic hours bore a significant relationship to any skill score measures (see Table 10). It is reasonable to conclude that higher cognitive ability leads to better prediction of a combination of rank order, elevation, and scatter.

Several personality dimensions had statistically significant relationships with skill score. At some point during or just after training, positive valence (r = -0.17, -0.18, and -0.24; $p < 0.05$; N = 142), conscientiousness (r = -0.19, -0.22; $p < 0.05$; N = 142), and conventionality (lack of openness; r = -0.22, -0.24; $p < 0.05$; N = 142) had a negative relationship to skill score (see Table 10). This evidence suggests that those who do not possess unusually high self-esteem, those who are not as thorough in their efforts, and those who are open-minded to new ideas will more accurately predict. Perhaps a very high level of self-esteem breeds overconfidence in the task. As previously discussed, thoroughness might result in refusal to accept less than perfect prediction, overthinking, and less accuracy. Furthermore, one who is open to new ideas might be better able to learn to master the prediction task.

Interests also demonstrated statistically significant relationships to a skill score metric that might reflect training feedback (a skill score measure other than *SS1*). Realistic interests (r = 0.17, 0.17, and 0.18; $p < 0.05$; N = 142), investigative interests ($r = 0.17$; $p < 0.05$; $N = 142$), and artistic interests ($r = 0.18$; $p < 0.05$; $N = 142$) were positively related to a skill score measure, while enterprising interests ($r = -0.18$; $p < 0.05$; $N = 142$) were negatively related to a skill score measure. In connection with training, more practical, problem-solving, creative, and less enterprising people better predict a combination of rank order, elevation, and scatter (see Table 10). This observation is not surprising.

In terms of prior experience making predictions of human behavior, the number of persons for whom predictions of human behavior were made in the past ($r = 0.19$; $p < 0.05$; $N = 142$) and the number of outcomes predicted in the past with a formal process ($r = 0.21$; $p < 0.05$; $N = 142$) bore a statistically significant positive relationship to skill score prior to any feedback training (i.e. for *SS1*; see Table 10). Experience can help, at least when subjects have not yet received any experimental training feedback. The number of outcomes for which predictions were made in the past based on formal training bore a statistically significant negative relationship to accuracy (r = -0.17; $p < 0.05$; N = 142; see Table 10). Perhaps past formal training instilled poor strategies.

Spring 2010 Results

Judgment criterion-related validity (r_a). The following individual difference variables had statistically significant and positive coefficients and improved model fit based on log likelihood chi-square tests: ACT composite score and ACT Mathematics score (see Table 20). Therefore, they added significantly to the prediction of within-person growth in clinical criterion-related validity over time. Given the intellectually challenging nature of the experiment, it was expected that cognitive ability (particularly mathematics ability) would be positively related to clinical judgment accuracy.

Based on between-subjects correlational relationships, there are several interesting findings. All sections of the ACT (verbal, mathematical, writing, reading comprehension, and scientific reasoning ability) as well as ACT composite score were significantly positively related to greater accuracy ($p < 0.05$) as measured by an r_a metric that possibly reflects experimental training (i.e., all measures of r_a except r_a1; see Table 19). Statistically significant correlations ranged from 0.21 to 0.32 (with Ns ranging from 85 to 100). The number of college academic hours for a subject was both positively and negatively related to an r_a measure that potentially reflected the receipt of feedback training (r = -0.19, 0.22; $p < 0.05$; N = 133; see Table 19). High school rank was positively related to such a metric (r = 0.24; $p < 0.05$; N = 133; see Table 19). Not surprisingly, those higher in cognitive ability predicted more accurately. Number of college academic hours reflects amount of exposure to college-level material and motivation. It is not clear why it would both positively and negatively relate to an r_a measure that might reflect training.

There were no statistically significant relationships for personality dimensions, but there were for social interests (r = -0.18; $p < 0.05$; N = 133; see Table 19). It is not surprising that those who are interested in interpersonal relationships would perform worse in a situation requiring highly impersonal assessment. It was expected that at least some personality dimensions (e.g., intellectance) as well as other interest dimensions (e.g., investigative) would have statistically significant relationships to an r_a measure. It is not clear why they were not found.

For experience, two variables had statistically significant negative relationships to an r_a metric: the number of outcomes for which a prediction was made in the past in a formal way (r = -0.18; N = 139) and the amount of formal statistics training received (r = -0.21, N = 140) (see Table 19). The latter variable, however, applied only to r_a prior to the receipt of training feedback in the Spring 2010 experiment (i.e., it was for block 1, r_a1). Thus, past experience making predictions might have instilled poor prediction strategies. The nature of formal statistical knowledge is unclear. If, for example, students had been sensitized to linear relationships only, then they might have been less likely to see nonlinear relationships. Therefore, they would have performed relatively poorly in block 1.

Skill score. The following individual difference variables had statistically significant and positive coefficients and improved model fit based on log likelihood chi-square tests: ACT composite score, ACT English score, ACT English Writing score, ACT Reading score, ACT Science Reasoning score, and high school rank (see Table 21). Those variables add significantly to the prediction of within-person growth in clinical accuracy over time. Given the cognitive challenge of the judgment task, it was expected that those higher in cognitive ability would achieve greater accuracy.

With regard to between-subjects relationships, all of these ACT scores as well as high school rank had statistically significant correlations ($p < 0.05$) with a skill score metric that potentially reflected experimental training. Such ACT correlations ranged from 0.20 to 0.39 (with *N*s ranging from 85 to 100; see Table 19). ACT Mathematics score was positively related in a statistically significant way to skill score metrics both related and unrelated to the receipt of feedback training (i.e. to *SS1* as well as to other skill score metrics). Thus, without any experimental training, people higher in mathematical ability tended to predict more accurately. High school rank was positively related to skill score metrics in a statistically significant way (r = 0.26, $p < 0.05$, N= 80). The data indicate that people higher in cognitive ability will predict more accurately.

Among personality characteristics, interests, and previous experience, only conventionality as measured by the IPC7 had a statistically significant relationship with a measure of skill score ($r = 0.20$, $p < 0.05$, $N = 133$; see Table 19). This is the only statistically significant relationship with a measure of accuracy that was positive in sign. In prior analyses, it was thought that conventionality (lack of openness to experience) would hinder predictive accuracy. However, in those instances, the skill score metric in question was potentially affected by training feedback (i.e., was not *SS1*). In this case, the skill score metric in question (*SS1*) was measured prior to subjects' receipt of any feedback. It is possible that lack of openness helps to prevent the consideration of irrelevant information when making a judgment. In the experiment, however, when a large amount of relevant information is present, conventionality (or lack thereof) may be less related to performance. In any case, it was surprising that other personality characteristics, interests, and prior experience had no statistically significant relationships with skill score.

Discussion

Except for findings regarding cognitive ability, the results are mixed. Depending on which dataset and measure of accuracy one is willing to adopt the findings from (Fall 2009 versus Spring 2010, r_a versus skill score), one might be inclined to ask people who are more agreeable and/or conventional (less open to experience) who have less formal statistics training, and who have rated many people in the past with a formal process to make predictions when training is unavailable and accuracy is paramount. When training is available and accuracy is important, one should teach those higher in cognitive ability, who possess investigative interests, and who are more open to experience to make judgments. Prior experience often has a negative impact on learning, possibly due to the inculcation of poor judgment strategies, so it might be best to start with blank slates.

The individual difference with the most robust relationship to accuracy (both as measured by r_a and skill score) was cognitive ability. Except for the analysis of which individual differences incremented prediction of change in accuracy for Fall 2009, cognitive ability (particularly as measured by the ACT and to a lesser extent as measured by high school rank and cumulative college GPA) was positively and pervasively related to more accurate prediction. Given the cognitively complex nature of predicting human behavior based on statistical information (especially a disordinal interaction), it is not surprising that cognitive ability played such an important role.

Among the personality dimensions, conventionality as measured by the IPC7 inventory had the most pervasive relationship to accuracy. Except for the between-subjects relationship to skill score in Spring 2010, the relationship of conventionality to accuracy was negative. As discussed, the positive relationship to skill score was for a metric of skill score that did not reflect any feedback training (i.e. *SS1*). Therefore, it appears that lack of openness is generally a detriment to accurate prediction when the experimental feedback training is experienced but not otherwise. As mentioned earlier, lack of openness might be a good thing if it filters out large amounts of irrelevant information one might normally receive, but if one is receiving a great deal of relevant information (as in experimental training), it might not be very significant. Positive valence, extroversion, and conscientiousness (all with negative between-subjects' relationships to accuracy) and intellectance and agreeableness (with positive between-subjects' relationships to accuracy) were occasionally of statistical significance.

For the RIASEC model, having investigative interests was, above all other interests, related to accuracy in a statistically significant and positive way (for both incrementing prediction of change in accuracy over time and explaining between-subjects' differences -- albeit limited to Fall 2009). It is expected that those who like learning and discovery will be more motivated to perform well in the experiments. To a statistically significant extent for Fall 2009 skill score, realistic interests both positively incremented prediction of accuracy over time and positively accounted for between-subjects' differences. Artistic interests positively explained between-subjects' differences in Fall 2009 skill score, too. Enterprising interests negatively accounted for between-subjects' differences in both r_a and skill score in a ststaistically significant manner for Fall 2009, and social interests negatively accounted for between-subjects r_a for Spring 2010 in a statistically significant way.

Among prior experiential variables, the number of outcomes predicted in the past in a formal manner had the highest frequency of statistically significant relationships with accuracy. For Fall 2009 analyses (both in terms of incrementing time's prediction of skill score and explaining between-subjects' differences in skill score), this variable was positive in sign. However, for the Spring 2010 between-subjects' analysis of r_a, it was negative in sign. If the past formal process instilled habits conducive to accurately predicting elevation and scatter but not rank order, then these results make sense. Length of prior training was negatively related to r_a in Fall 2009 for incrementing prediction of accuracy over time as well as for explaining between-subjects' differences in r_a. With regard to between-subjects' differences, the number of persons for whom predictions were made in the past was positively related to accuracy (for Fall 2009 skill score). With regard to between-subjects' differences, the number of outcomes for which predictions were made in the past based on formal training (for Fall 2009 skill score), the amount of formal education in judgment and decision making (for Fall 2009 r_a), and the amount of formal education in statistics (for Spring 2010 r_a) were negatively related to accuracy. To reiterate, past experience can be both good and bad for future accuracy.

Insight

The sixth research question concerns when and to what extent (if any) individuals possess insight about the optimal strategy for making accurate judgments. As discussed, determination of when and whether insight was achieved was based upon narrative self-reports later separately coded by this author and another Ph.D. candidate in industrial-organizational psychology. For both Fall 2009 and Spring 2010, there were a fair amount of missing data, and due to relative consistency across blocks of the subjects' binary-coded self-reports, there was generally not much within-person variability within each insight dimension across blocks. Therefore, the dimensions for insight (awareness, use, and correct use of the disordinal interaction) were not treated as within-person variables. Instead, each dimension was measured as a set of frequencies or as a between-subjects variable based on the percentage of blocks for which each subject was coded "yes" for the dimension. The between-subjects variables were normalized via an arcsine transformation described in Table 5.[67]

Fall 2009 Results

In Table 13 appears the initial interrater agreement between the coders. That interrater agreement was very high (no lower than 82% in any situation, with a mean of 93%) and was later reconciled after discussion.

On a conceptual level, one would expect self-reports of awareness, use, and correct use of the disordinal interaction to be related. In fact, they typically would be nested, such that those who indicate that they correctly used the disordinal interaction would also indicate that they used the disordinal interaction, and those who indicate that they used the disordinal interaction would also indicate that they were aware of the disordinal interaction. Evidence of relatedness of insight

[67] These transformed variables are named "TrPerExprAware", "TrPerSaidUsed", and "TrPerUsedCorrectly". "TrPerSaidUsed3rdVar" represents an unanalyzed variable representing the arcsine transformed percentage of blocks for which the subject said that s/he used information other than cognitive ability, interesting/boring expectation for the job, or the interaction between them. Amazingly, some subjects said with complete sincerity that they used information that was unavailable to them – except perhaps in their imaginations – to try to make predictions.)

dimensions appears as moderate to strong positive and statistically significant ($p < 0.05$) intercorrelations in Table 10 ($r_{\text{Aware, Used}} = 0.88$ with $N = 134$; $r_{\text{Used, Used Correctly}} = 0.39$ with $N = 129$; $r_{\text{Aware, Used Correctly}} = 0.41$ with $N = 129$).[68]

Oddly enough, however, nesting did not occur in many cases. For instance, some subjects circled "Yes" in response to the question about whether they used the disordinal interaction, but at the same time the subjects' narrative responses indicated that they thought that the relationship between cognitive ability and performance was based solely on a single regression line. It is possible that subjects nevertheless robotically and accurately read from scatterplot feedback to later predict performance based on the moderator of how interesting or boring the applicant was expected to find the job. However, accurately using procedures is not synonymous with conceptual understanding (i.e. insight). Why does this matter from a practical perspective? Insight makes replication of accurate performance more likely, especially when new tasks are ambiguous.

To what extent and how were insight dimension ratings related to objective measures of judgment accuracy (i.e. r_a and skill score)? Although ratings of the percentage of blocks for which the disordinal interaction was used correctly had a statistically significant (and positive) coefficient in a multilevel regression model where time was the independent variable predicting changes in r_a, they did not improve fit over a multilevel regression model without those ratings included (see Table 14). When skill score was the dependent variable, model fit was improved, but the coefficients for the insight variables were not statistically significant (see Table 15). In

[68] These observed correlations (and later ones) can be corrected for dichotomization. Dichotomization artificially reduces variability of the dichotomized variable(s), which reduces the magnitude of a correlation that is based on such dichotomized variable(s). Thus, the correlations would be greater in magnitude after correction. Insight dimensions can be thought of as truly continuous variables that have been artificially split due to difficulties in interpreting subjects' narrative self-reports. For example, after corrections are made, $r_{\text{Used, Used Correctly}}$ increases from 0.39 to 0.78, and $r_{\text{Aware, Used Correctly}}$ increases from 0.41 to 0.82. $r_{\text{Aware, Used}}$ becomes a perfect 1.00. The splits for these corrections come from the percentages for the insight dimensions in Table 16 (for Fall 2009). More specifically, they are based on the percentages of those who achieved insight (100% - the percentage for "Never") after interpreting insight as being indeterminate for a subject if any reversals exist. This was thought to be a reasonable compromise position, because ignoring reversals leads to a larger percentage of achievers, and interpreting reversals as indicating failure to obtain insight leads to a smaller percentage of achievers. See Hunter and Schmidt (2004, p. 36) for the procedure for making corrections for dichotomization.

other words, insight did not significantly explain change in judgment accuracy beyond what is explained by the passage of time. With regard to between-subjects analysis, the only statistically significant correlation for either r_a or skill score was the positive one between the percentage of blocks for which the disordinal interaction was used properly (based on ratings) and overall r_a ($r = 0.18$, $p < 0.05$, $N = 129$; see Table 10).[69] Thus, there is limited evidence that insight is related to actual performance for the Fall 2009 data.

Time to insight is very difficult to determine partly because of the presence of reversals in the coding for self-reports. As discussed earlier, simply using objective measures of accuracy (r_a and skill score) is insufficient to ascertain insight, because some subjects may have simply followed instructions, rotely applied feedback, or predicted accurately due to random chance. In Table 16 appear tables showing time to insight based on each of the three insight dimensions that were coded. Each of the tables measures insight in 3 different ways, because it is debatable how reversals should be treated. One can ignore them, consider their existence as indicating that insight is indeterminate, or consider their existence as indicating that insight was not achieved. The only situations in which a plurality of subjects achieves insight is for mere awareness of the disordinal interaction and use of the disordinal interaction when reversals are ignored. Otherwise, the plurality never achieves insight. In fact, the vast majority of subjects never achieves insight as measured by ratings of self-reported correct use of the disordinal interaction. Generally, if insight is achieved in any way, it is usually achieved immediately after the first block of predictions. Given that subjects had not received any feedback before self-reporting any insight for the first block, feedback does not seem to be very conducive to promoting insight.

[69] As was discussed in a previous footnote for other insight correlations, this observed correlation can be corrected for dichotomization. After the correction is made, it equals 0.23.

Spring 2010 Results

In Table 22 appears the initial interrater agreement between the coders. That interrater agreement was high (no lower than 66% for any situation, with a mean of 84%) and was later reconciled after discussion.

Evidence of relatedness of insight dimensions appears as moderate to strong positive and statistically significant ($p < 0.05$) intercorrelations in Table 19 ($r_{\text{Aware, Used}} = 0.86$ with $N = 135$; $r_{\text{Used, Used Correctly}} = 0.59$ with $N = 113$; $r_{\text{Aware, Used Correctly}} = 0.58$ with $N = 113$).[70] Again, nesting of insight dimensions did not occur in many cases.

To what extent and how were insight dimension ratings related to objective measures of judgment accuracy (i.e. r_a and skill score)? For both r_a and skill score as the dependent variable, all insight dimensions had a statistically significant (and positive coefficient) in a multilevel regression model where time was the independent variable, and they all improved fit over a multilevel regression model without those ratings included (see Tables 23 and 24). In other words, insight significantly explained change in judgment accuracy beyond what is explained by the passage of time. With regard to between-subjects analysis, statistically significant correlations with r_a measures that potentially reflect the impact of training were as follows: awareness of the disordinal interaction with $r = 0.23$, 0.24, and 0.25 where $N = 135$); usage of the disordinal interaction with $r = 0.29$, 0.21, and 0.28 when $N = 135$; and correct usage of the disordinal interaction with $r = 0.30$, 0.28, and 0.33 when $N = 113$.[71] Statistically significant correlations with skill score measures that potentially reflect the impact of training were as follows: awareness of the disordinal interaction with $r = 0.24$, 0.26, and 0.26 where $N = 135$); usage of the disordinal interaction with $r = 0.32$, 0.32, and 0.21 when $N = 135$; and correct usage

[70] As was discussed in a previous footnote for other insight correlations, these observed correlations can be corrected for dichotomization. After the corrections are made, they all equal 1. The splits for these corrections come from the percentages for the insight dimensions in Table 25 (for Spring 2010).

[71] As was discussed in a previous footnote for other insight correlations, these observed correlations can be corrected for dichotomization. After the corrections are made, they equal 0.29, 0.30, and 0.31, respectively, for Awareness; 0.36, 0.26, and 0.35, respectively, for Use; and 0.48, 0.45, and 0.52, respectively, for Correct Use.

of the disordinal interaction with $r = 0.26$, 0.26, and 0.24 when $N = 113$. See Table 19.[72] There were no statistically significant correlations for r_a or skill score for prediction before any task feedback was provided. Thus, there is substantial evidence that insight is related to actual performance for the Spring 2010 data.

Again, time to insight is very difficult to determine partly because of the presence of reversals in the coding for self-reports. Table 25 shows time to insight based on each of the three insight dimensions that were coded. Each of the tables measures insight in 3 different ways, because it is debatable how reversals should be treated. In comparison to Fall 2009, insight is achieved far more often in Spring 2010 (compare Table 16 to Table 25). This may be due to the stronger disordinal interaction in the Spring 2010 data, the provision to all feedback groups in Spring 2010 but not Fall 2009 of scatterplots showing the relationship of cognitive ability to performance for interesting jobs and then separately for boring jobs, and/or some other factors. In any case, insight across all three dimensions is usually achieved at some point. As time progresses, it is less likely to be achieved, so subjects generally achieve insight early if ever (see Table 25). Unlike in Fall 2009, those who achieve insight in Spring 2010 generally do so after receiving feedback that follows the first block. Thus, it appears that in Spring 2010 (but not so much in Fall 2009) feedback helps many people achieve insight.

[72] As was discussed in a previous footnote for other insight correlations, these observed correlations can be corrected for dichotomization. After the corrections are made, they equal 0.30, 0.33, and 0.33, respectively, for Awareness; 0.40, 0.40, and 0.26, respectively, for Use; and 0.41, 0.41, and 0.38, respectively, for Correct Use.

GENERAL DISCUSSION

There were a number of interesting findings. Only for predicting rank order (as opposed to elevation or scatter) could assessors learn to outperform a mechanical approach in the applicant selection context. This ability was greater when the disordinal interaction was stronger and all subjects received at least some feedback about the disordinal interaction (in Spring 2010). Results were not statistically significant, however. Furthermore, the lack of psychological realism of the experiments puts into question whether the same subjects would be able to perform comparably well if there were no experimenter providing task information feedback. As already discussed, real-world judges usually must see and use functional forms without the level of guidance provided in the experiments. In addition, that any helpful nonlinear functional forms even exist in naturalistic applicant selection contexts is uncertain.

Nevertheless, to some extent assessors *can* learn to improve their judgment accuracy, although the evidence is somewhat contradictory and complicated by statistically insignificant results. In both experiments, this improvement held true for predicting a combination of rank order, elevation, and scatter (i.e. for skill score). There is some statistically significant evidence that for skill score those who receive express feedback about the disordinal interaction become more accurate while those who do not become less accurate (group 1 versus group 3 in Fall 2009). On the other hand, for predicting only rank order the evidence indicates that performance declined generally (in Fall 2009) or increased generally and for all feedback groups considered separately (in Spring 2010). Of this incongruous evidence, the only data that were statistically significant were for those groups who received explicit wording that the disordinal interaction might or did exist and improved in accuracy over time (groups 1 and 2 in Spring 2010).

In general, the determinants of r_a are consistent with what one would expect for changes in r_a in the experiments. In Fall 2009, overall r_a declined over time while overall C, r_z, G, and C_{xj} declined and overall R_s increased over time. (As is the common theme, the confidence intervals of many all slope trajectories contained both positive and negative values, so conclusions are again ambiguous). The overall decrease in use of unmodeled knowledge (*C*) over time detracted from accuracy, as did the overall decrease in criterion-relevant part of unmodeled knowledge (r_z). The overall waning of the relative weight given to the disordinal interaction (C_{xj}) is consistent with overall decreases in *C* and r_z. Although it would not have permitted the judges to outperform the mechanical model even if its growth trajectory were positive, the overall decrease in mechanical knowledge (*G*) over time negatively impacted r_a. The overall increase in cognitive control (R_s) over time had a mixed impact. It decreased the unmodeled component of r_a but increased the mechanical component of it. On the whole, the overall change in R_s over time did not improve accuracy.

The determinants of r_a in Spring 2010 also tell a consistent story (the issue of statistical insignificance aside). As overall r_a (and skill score) increased over time, so too did overall use of unmodeled knowledge (*C*), overall use of the criterion-relevant part of unmodeled knowledge (r_z), and the overall relative weight given to the disordinal interaction (C_{xj}). It appears that subjects generally were able to increasingly employ configural reasoning (i.e. leverage information about the disordinal interaction) to improve their accuracy. Overall use of mechanical knowledge (*G*) and cognitive control (R_s) fell over time. Perhaps subjects' change of focus away from the linear mechanical approach and toward unmodeled information explains the change in overall use of mechanical knowledge. Declining cognitive control would have contributed to the unmodeled component of r_a, which is consistent with the other trends. It is possible that these trends can be explained by the stronger interaction in Spring 2010 as well as the provision of information to all subjects in Spring 2010 of some criterion-relevant information about the disordinal interaction.

There is also some evidence that judges *can* be made less confident in their judgment strategies. Although the evidence for changes in confidence was less contradictory than was the evidence for changes in accuracy, lack of statistical significance once again complicates the ability to draw conclusions. This is especially true for confidence as measured after every block (as opposed to overall). Aside from results for changes in relative confidence in Fall 2009, findings for confidence measured after every block were for the most part not statistically significant. For both experiments (but especially for Fall 2009), there is some evidence that subjects who receive more information about the disordinal interaction decline in confidence and those who receive less increase in it. For measuring overall confidence, one sees only declines for subjects. Declines usually were not statistically significant for group 1, which received the least amount of feedback. Those findings for overall confidence are at least partly consistent with findings for confidence measured after every block.

Since findings for changes in accuracy and to a lesser extent changes in confidence were often statistically insignificant, it is difficult to argue that there is a statistically significant relationship between changes in accuracy and changes in confidence. The only expected cross-over patterns (i.e., declining confidence with increasing accuracy for groups that received more feedback information about the disordinal interaction, and vice-versa) were for skill score and both absolute and relative confidence measured after every block, all in Fall 2009. Groups 2 and 3 (receiving explicit information about the disordinal interaction) increased in accuracy but declined in confidence, while group 1 (receiving no explicit information about the disordinal interaction) declined in accuracy but increased in confidence. Otherwise, the findings are not consistent with the idea that criterion-related task information feedback leads to increased accuracy and decreased confidence while lack of it leads to decreased accuracy and increased confidence.

The several findings for individual differences are discussed above and thus are only summarily reviewed. Without experimental training (i.e. with regard to relationships with r_a or skill score immediately after block 1, before any task information feedback was received), the following are the characteristics of those who should be selected to make judgments, especially for predicting rank order only (because of relationships to larger r_a values): people who are more agreeable (based on Fall 2009 data) and those who have not received much (if any) formal statistics education (based on Spring 2010 data). If that former education promoted linear reasoning primarily (as may be the case in most introductory statistics classes that spend more time on correlation, simple linear regression, and main effects than interactions or other nonlinear forms), then that outcome makes sense when configural reasoning is the key to maximizing accuracy. Those who have more experience making past predictions of human behavior in a formal way (based on Fall 2009 data), who have rated more people in the past (based on Fall 2009 data), who have greater mathematical ability (based on Spring 2010 data), and who are more closed to new experience (based on Spring 2010 data) are better at predicting a combination of rank order, elevation, and scatter (i.e. achieve higher skill scores). Being closed to new experience can be beneficial to accurate prediction if information from the environment is misleading or distracting or overloads working memory. Please note that the following data are based on between-subjects correlations.

Based on their relationship to accuracy after time is considered as well as based on between-subjects relationships after feedback is received, the major characteristics of those expected to be most accurate are those higher in cognitive ability, more open to new experience, and more investigative. The number of outcomes for which predictions were made in the past was sometimes positively related to accuracy and sometimes negatively related to it. One needs to know more about this type of experience before knowing how to use it as a predictor. If past judgment was based on faulty assumptions, then it may promote poor judgment in the future.

The findings about the ability of individuals to attain insight about the optimal strategy for making accurate judgments are both encouraging and discouraging. For example, insight usually was never achieved in Fall 2009, whereas it usually was achieved in Spring 2010. That the insight dimensions were nested in theory but not in their operationalization is interesting but makes measurement of insight more complicated. At least the dimensions showed nontrivially positive correlations with each other. So too does the presence of “reversals” in responses to questions about insight. It is reassuring that at the very least coders can achieve relatively high agreement on whether others have achieved insight. It appears that when the disordinal interaction is stronger and more subjects receive at least some feedback about the disordinal interaction (in Spring 2010 compared to Fall 2009), insight is more often achieved, is generally achieved sooner, is more likely to explain change in accuracy beyond the effect of the passage of time, and is more likely to have a statistically significant relationship to objective measures of accuracy (r_a and skill score).

In general, the studies did not unambiguously answer the research questions. Overlapping confidence intervals and *p*-values slightly greater than 0.05 suggest that larger samples are needed to draw more meaningful conclusions. These phenomena even affected the decision about how to undertake analyses. For example, the analyses that looked at whether individual differences explained variability in accuracy beyond what was explained by the passage of time did not look at the possibility that time and the individual difference interacted. In other words, the individual difference variables were entered into equations as main effects only. Given the frequently overlapping confidence intervals for slope trajectories among the feedback groups as well as periodic inability to find statistical significance for a time x condition interaction, it was felt that an examination of time x individual difference interactions (or time x condition x individual difference interactions) would add complexity but not yield statistically significant results.

Lack of statistical significance of some of the interactions may indicate that the study design should be revised once again. In particular, it might be helpful to retain the stronger disordinal interaction in Spring 2010 but the greater feedback differences among conditions in Fall 2009. The stronger the disordinal interaction, the more likely it is to be noticed. Providing absolutely no explicit information in words or displays about the disordinal interaction to one of the feedback groups (as with group 1 in Fall 2009) may be necessary to determine the extent to which subjects can see the disordinal interaction in the data.

Another challenge is generalizability of findings. Ideally, the subjects would not be a convenience sample of mostly undergraduates but rather those who regularly make selection decisions for the workplace and academic environments (i.e. hiring managers and admissions officers). What is meaningful to an experienced hiring manager or admissions officer potentially educated in selection matters might not be meaningful (at least in the same way) to an undergraduate at a large state university (and vice-versa). In other words, the ideal experimental design for the convenience sample (particularly the ideal nature of feedback provided) might not be the ideal experimental design for the population of real interest. While a convenience sample might provide interesting findings that might inform real-world contexts to some extent, whenever possible it is important to recruit from the population of interest rather than a convenience sample.

There is a serious dearth of research using Lens Model parameters or skill scores to study training to improve clinical accuracy under at least quasi-realistic environmental and feedback conditions. For this purpose, there does not appear to be a paradigm to serve as an adequate substitute for the Lens Model or skill score either. Also acutely lacking is research on individual differences that explain which judges predict and learn to predict most accurately. Given the ambiguity of many of the study results, the recurring meta-analytic and literature review findings that mechanical methods systematically outperform holistic approaches, the potentially grave consequences of inaccuracy in predicting work and academic outcomes, and the costs of training, there is a need for conducting additional primary studies with populations of interest rather than convenience, larger samples of subjects, additional blocks of prediction, and improved environmental and feedback designs. Perhaps future research will also find other informative ways of objectively measuring predictive accuracy and the determinants of that accuracy. Although this dissertation covered a wide range of individual differences, there are additional individual differences that might explain variability among people in accuracy and which thus should be explored (e.g., global self-efficacy). Furthermore, new investigations of how best to conceptualize and measure insight as well as insight's relationship to actual accuracy are crucial to making training more effective. Whatever its limitations, it is hoped that this dissertation has at least raised in the minds of its readers substantive and methodological concerns that motivate and inform future exploration of what the author considers to be intriguing and consequential issues.

REFERENCES

Adams, J. S. (1963). Toward an understanding of inequity. *Journal of Abnormal and Social Psychology, 67*, 422-436.

Adams, J. S. (1965). Inequity in social exchange. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 267-296). New York: Academic Press.

Adelman, L. (1981). The influence of formal, substantive, and contextual task properties on the relative effectiveness of different forms of feedback in multiple-cue probability learning tasks. *Organizational Behavior and Human Performance*, 27, 423-442.

Ader, H. J., & Mellenbergh, G. J. (Eds.). (1999). *Research methodology in the life, behavioural and social sciences*. London: Sage.

Archer, J. (1988). The sociobiology of bereavement: A reply to Littlefield and Rushton. *Journal of Personality and Social Psychology, 55*, 272-278.

Arkes, H. R. (1991). Costs and benefits of judgment errors: Implications of debiasing. *Psychological Bulletin, 110*(3), 486-498.

Arkes, H. R., & Blumer, C. (1985). The psychology of sunk cost. *Organizational Behavior and Human Decision Processes, 35*, 125-140.

Armelius, B. A., & Armelius, K. (1974). The use of redundancy in multiple-cue judgments: Data from a suppressor-variable task. *American Journal of Psychology*, 87 (3), 385-392.

Ashton, A. H. (1982). An empirical study of budget-related predictions of corporate executives. *Journal of Accounting Research*, 20 (No. 2 Pt I), 440-449.

Balzer, W. K., Doherty, M. E., & O'Connor, R. (1989). Effects of cognitive feedback on performance. *Psychological Bulletin*, 3, 410-433.

Balzer, W. K., Sulsky, L. M., Hammer, L. B., & Sumner, K. E. (1992). Task information, cognitive information, of functional validity information: Which components of cognitive feedback affect performance. *Organizational Behavior and Human Decision Processes*, *53*, 35-54.

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*, 1-26.

Björkman M. (1972). Feedforward and feedback as determiners of knowledge and policy: Notes on a neglected issue. *Scandinavian Journal of Psychology*, *13,* 152–158.

Bliese, P. D. (2000). Within-group agreement, non-independence, and reliability. In K.J. Klein & S.W.J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations* (pp. 349-381). San Francisco, CA: Jossey Bass.

Bowman, E. H. (1963). Consistency and optimality in managerial decision making. *Management Science*, 9, 310-321.

Brannick, M. T., & Brannick, J. P. (1989). Nonlinear and noncompensatory processes in performance evaluation. *Organizational Behavior and Human Decision Processes*, 44, 97-122.

Brehmer, B. (1974). The effect of cue intercorrelation on interpersonal learning of probabilistic inference tasks. *Organizational Behavior and Human Performance, 12*, 397–412.

Brehmer, B. (1979). Preliminaries to a psychology of inference. *Scandinavian Journal of Psychology, 20*, 193-210.

Brehmer, B. (1980). In one word: Not from experience. *Acta Psychologica*, 45, 223-241.

Brehmer, A., & Brehmer, B. (1988). What have we learned about human judgment from thirty years of policy capturing? In B. Brehmer & C.R.B. Joyce (Eds.), *Human judgment: The SJT view* (pp. 75-114). Amsterdam: North-Holland.

Brehmer, B., & Joyce, C. R. B. (Eds.) (1988). *Human judgment: The SJT view*. Amsterdam: North-Holland.

Brehmer, B., & Kuylenstierna, J. (1980). Content and consistency in probabilistic inference tasks. *Organizational Behavior and Human Performance*, 26, 54-64.

Brogden, H. E. (1949). When testing pays off. *Personnel Psychology, 2,* 171-183.

Brunswik, E. (1952). *The conceptual framework of psychology*. Chicago: University of Chicago Press.

Bryk, A. S., & Raudenbush, S. W. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Budescu, D. V. (1993). Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression. *Psychological Bulletin, 114*, 542-551.

Camerer, C. F. (1981a). *The validity and utility of expert judgment*. Unpublished doctoral dissertation, University of Chicago Graduate School of Business, Retrieved February 27, 2009, from ProQuest databases.

Camerer, C. F. (1981b). General conditions for the success of bootstrapping models. *Organizational Behavior and Human Performance, 27*, 411-422

Camerer, C. F., & Johnson, E. J. (1991). The process-performance paradox in expert judgment: How can experts know so much and predict so badly? In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: prospects and limits* (pp. 195-217). New York, NY: Cambridge University Press.

Campbell, J. P. (1974). *A Monte Carlo approach to some problems inherent in multivariate prediction: With special reference to multiple regression.* Technical Report # 2002. Personnel and Training Research Programs. Office of Naval Research.

Campbell, J. P. (1990). Modeling the performance prediction problem in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 1, pp. 687-732). Palo Alto, CA: Consulting Psychologists Press.

Castellan, N. (1972). The analysis of multiple criteria in multiple cue judgment tasks. *Organizational Behavior and Human Decision Processes*, 8, 242-261.

Castellan, N. J., Jr. (1992). Relations between linear models: Implications for the lens model. *Organizational Behavior and Human Decision Processes*, 51, 364-381.

Chapman, N.J. & Chapman, J.P. (1971). Test results are what you think they are. *Psychology Today*, 5, 18-22, 106-110.

Chi, M.T.H. (2006). Two approaches to the study of experts' characteristics. In K.A. Ericsson, N. Charness, P.J. Feltovich & R.R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 21-30). New York, NY: Cambridge University Press.

Cohen, B. H. (2001). *Explaining psychological statistics* (2nd ed.). New York, NY: Wiley.

Cooksey, R. W. (1996). *Judgment analysis: Theory, methods, and applications*. New York, NY: Academic Press.

Cooksey, R. W., & Freebody P. (1985). Generalized multivariate lens model analysis for complex human inference tasks. *Organizational Behavior and Human Decision Processes*, 35, 46-72.

Cooper, R. P., & Werner, P. D. (1990). Predicting violence in newly admitted inmates: A lens model analysis of staff decision making. *Criminal Justice and Behavior*, 17 (4), 431-447.

Cowles, M., & Davis, C. (1982). On the origins of the .05 level of statistical significance. *American Psychologist, 37*, 553-558

Crandall, B., Klein, G., & Hoffman, R. R. (2006). *Working minds: A practitioner's guide to cognitive task analysis.* Cambridge, MA: MIT Press.

Cronbach, L. J., & Gleser, G. C. (1953). Assessing similarity between profiles. *Psychological Bulletin, 50*, 456-473.

Cronbach, L. J., & Gleser, G. C. (1965). *Psychological tests and personnel decisions*. Urbana, IL: University of Illinois Press.

Dalal, R.S., Bonaccio, S., Highhouse, S., Ilgen, D.R., Mohammed, S., & Slaughter, J.E. (2010). What if industrial-organizational psychology decided to take workplace decisions seriously? *Industrial and Organizational Psychology: Perspectives on Science and Practice, 3*(4), 2-46.

Dawes, R. M. (1964). *Toward a general framework for evaluation*. Ann Arbor: University of Michigan,

Department of Psychology.

Dawes, R. M. (1971). A case study of graduate admissions: Applications of three principles of human decision making. *American Psychologist*, 26, 180-188.

Dawes, R. M. (1976). *Shallow psychology*. In J. S. Carroll & J. W. Payne (Eds.), Cognition and social behavior. Hillsdale, NJ: Erlbaum.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, 34(7), 571-582

Dawes, R., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin*, 81, 95–106.

Deane, D. H., Hammond, K. R., & Summers, D. A. (1972). Acquisition and application of knowledge in complex inference tasks. *Journal of Experimental Psychology*, 92 (1), 20-26.

Deci, E. L. (1971). Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology, 18,* 105-115.

Deci, E. L. (1972a). The effects of contingent and non-contingent rewards and controls on intrinsic motivation. *Organizational Behavior and Human Performance, 8,* 217-229.

Deci, E. L. (1972b). Intrinsic motivation, extrinsic motivation, and inequity. *Journal of Personality and Social Psychology, 22*, 113-120.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum.

Deese, J., & Kaufman, R. A. (1957). Serial effects in recall of unorganized and sequentially organized verbal material. *Journal of Experimental Psychology, 54*(3):180-187.

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7, 177–188.

Doherty, M. E. (1993). A laboratory scientist's view of naturalistic decision making. In G. Klein, R. Calderwood, J. Oransu, & C.E. Zsambok (Eds.), *Decision making in action: Models and methods*. Norwood, NJ: Ablex Publishing, 362-388.

Doherty, M. E., & Balzer, W. K. (1988). Cognitive feedback. In B. Brehmer & C.R.B. Joyce (Eds.), *Human judgment: The SJT view* (pp. 163-197). Amsterdam: North-Holland.

Dougherty, T. W., Ebert, R. J., & Callender, J. C. (1986). Policy capturing in the employment interview. *Journal of Applied Psychology*, 71, 9-15.

Drory, A. (1982). Individual differences in boredom proneness and task effectiveness at work. *Personnel Psvchologv*, 35, 141-151.

Dudycha, L. W., & Naylor, J. C. (1966). Characteristics of the human inference process in complex choice behavior situations. *Organizational Behavior and Human Performance*, 1, 110-128.

Edwards, W. (1975). Comment. *Journal of the American Statistical Association, 70*, 291-293.

Einhorn, H. J. (1970). The use of nonlinear, noncompensatory models in decision making. *Psychometrika*, *73*(3), 221-230.

Einhorn, H. J. (1971). Use of nonlinear, noncompensatory models as a function of task and amount of information. *Organizational Behavior and Human Performance*, 6, l-27.

Einhorn, H. J. (1974). Cue definition and residual judgment. *Organizational Behavior and Human Performance*, 12(1), 30-49.

Einhorn, H. J. (1986). Accepting error to make less error. *Journal of Personality Assessment*, 50(3), 387-395.

Einhorn, H. J., & Hogarth, R. M. (1975). Unit weighting schemes for decision making. *Organizational Behavior and Human Performance*, 13, 171-192.

Einhorn, H. J., & Hogarth, R. M. (1978). Confidence in judgment: Persistence of the illusion of validity. *Psychological Review*, 85(5), 395-416.

Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgment and choice. *Annual Review of Psychology, 32*, 53-88.

Einhorn, H.J., Kleinmuntz, D.N., & Kleinmuntz, B. (1979). Linear regression and process-tracing models of judgment. *Psychological Review*, *86*, 465-485.

Elrod, T., Johnson, R. D., & White, J. (2004). A new integrated model of noncompensatory and compensatory decision strategies. *Organizational Behavior and Human Decision Processes*, *95*(1), 1-19.

Festinger, L. (1957). *A theory of cognitive dissonance.* Stanford, CA: Stanford University Press.

Fischhoff, B. (1975): Hindsight / foresight: The effect of outcome knowledge on judgement under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, *1*, 288-299.

Fischhoff, B., Slovic, P., & Lichtenstein. S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology - Human Perception & Performance, 3*, 552-564.

Fisher, C. (1991). *Boredom at work: a neglected concept*. Bond University School of Business Discussion Paper No. 19.

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples of an indefinitely large population. *Biometrika, 10*, 507–521.

Fitzmaurice, G. M., Laird, N.M., & Ware, J. H. (2004). *Applied longitudinal analysis*. Hoboken, NJ: John Wiley & Sons, Inc.

Foderado, L. W. (2009, November 1). The whole applicant. *New York Times*. Retrieved from http://www.nytimes.com/2009/11/01/education/edlife/01admission-t.html

Galbraith, J. T. (1984). *Training assessment center assessors: Applying principles of human judgment*. Unpublished doctoral dissertation, Bowling Green State University, Bowling Green, OH.

Ganzach, Y. (1993). Goals as determinants of nonlinear noncompensatory judgment strategies: Leniency vs strictness. *Organizational Behavior and Human Decision Processes*, *56*, 422-440.

Ganzach, Y. (1994). Theory and configurality in expert and layperson judgment. *Journal of Applied Psychology*, *79*(3), 439-448.

Ganzach, Y. (1998). Nonlinear models in decision making: The diagnosis of psychosis versus neurosis from the MMPI. *Organizational Behavior and Human Decision Processes, 74*(1), 53–61.

Ganzach, Y., Kluger, A. N., & Klayman, N. (2000). Making decisions from an interview: Expert measurement and mechanical combination. *Personnel Psychology, 53*, 1-20.

Ghiselli, E. E., Campbell, J. P., & Zedeck, S. (1981). *Measurement theory for the behavioral sciences*. San Francisco, CA: Freeman.

Goldberg, L. R. (1970). Man versus model of man: A rationale, plus some evidence, for a method of improving on clinical inferences. *Psychological Bulletin*, 73 (6), 422- 432.

Goldberg, L. R. (1971). Five models of clinical judgment: an empirical comparison between linear and nonlinear representations of the human inference process. *Organizational Behavior and Human Performance, 6*, 458-479.

Goldberg, L. R. (1990). An alternative "Description of Personality": The Big-Five factor structure. *Journal of Personality and Social Psychology, 59*(6), 1216-1229.

Gratz v. Bollinger, 539 U.S. 244 (2003).

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Grether, D. M., & Plott, C. R. (1979). Economic theory of choice and the preference reversal phenomenon. *American Economic Review, 69*, 623-38.

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The Clinical–Statistical Controversy, *Psychology, Public Policy, and Law*, 2, 293–323.

Grove, W. M., Zald, D., H., Lebow, B., S., Snitz, B., E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, 12(1), 19-30.

Grutter v. Bollinger, 539 U.S. 306 (2003).

Guastello, S.J. (1995). *Chaos, catastrophe, and human affairs: Applications of nonlinear dynamics to work, organizations, and social evolution*. Mahwah, NJ: Lawrence Erlbaum.

Guastello, S. J. (2002). *Managing emergent phenomena: Nonlinear dynamics in work organizations*. Mahwah, NJ: Lawrence Erlbaum Associates.

Guastello, S.J., Koopmans, M., & Pincus, D. (Eds.). (2009). *Chaos and Complexity in Psychology: Theory of Nonlinear Dynamics*. NY: Cambridge University Press.

Guion, R. M. (1991). Personnel assessment, selection, and placement. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 2, pp. 327-397). Palo Alto, CA: Consulting Psychologists Press.

Guion, R. M. (1998). *Assessment, measurement, and prediction for personnel decisions*. Mahwah, NJ: Lawrence Erlbaum.

Hammond, K. R. (1955). Probabilistic functionalism and the clinical method. *Psychological Review*, 62, 255-262.

Hammond, K. R. (1971). Computer graphics as an aid to learning. *Science, 172*, 903-908.

Hammond, K. R. (1978). *Toward increasing competence of thought in public policy formation*. In K. R. Hammond (Ed.), Judgment and decision in public policy formation. Denver, CO: Westview Press.

Hammond, K. R. (1986). *A theoretically based review of theory and research in judgment and decision making* (Report No. 260). Center for Research on Judgment and Policy, University of Colorado, Boulder, CO.

Hammond, K. R. (1988). Judgment and decision making in dynamic tasks. *Information and Decision Technologies*, *14*, 3-14.

Hammond, K. R. (2007). *Beyond rationality: The search for wisdom in a troubled time*. New York, NY: Oxford University Press.

Hammond, K. R., Hursch, C. J., & Todd, F. J. (1964). Analyzing the components of clinical inference. *Psychological Review, 71,* 438–456.

Hammond, K. R., McClelland, G., & Mumpower, J. (1980). *Human judgment and decision making: Theories, methods, and procedures*. New York: Praeger Scientific.

Hammond, K. R., & Summers, D. A. (1965). Cognitive dependence on linear and nonlinear cues. *Psychological Review, 72,* 215–224.

Hammond, K. R., & Summers, D. A. (1972). Cognitive control. *Psychological Review, 79,* 58-67.

Hammond, K. R., Summers, D. A., & Deane, D. H. (1973). Negative effects of outcome-feedback in multiple-cue probability learning. *Organizational Behavior and Human Performance*, *9*, 30-34.

Hastie, R., & Dawes, R.M. (2001). *Rational choice in an uncertain world*. Thousand Oaks, CA: Sage Publications.

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21, 1539–1558.

Highhouse, S. (2008). Stubborn reliance on intuition and subjectivity in employee selection. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *1*, 333-342.

Hill, A. B. (1975). Work variety and individual differences in occupational boredom. *Journal of Applied Psychology*, 60(1), 128-131.

Hoffman, P. J. (1960). The paramorphic representation of clinical judgment. *Psychological Bulletin*, *57*, 116-131.

Hoffman, P. J., Earle, T. C., & Slovic, P. (1981). Multidimensional functional learning (MFL) and some new conceptualizations of feedback. *Organizational Behavior and Human Performance*, *27,* 75-102.

Hogarth, R. M. (1975). Cognitive processes and the assessment of subjective probability distributions. *Journal of the American Statistical Association*, *70*, 271-289.

Hogarth, R. M. (1981). Beyond discrete biases: Functional and dysfunctional aspects of judgmental heuristics. *Psychology Bulletin, 90*, 197-217.

Hogarth, R. M. (2001). *Educating intuition.* Chicago: University of Chicago Press.

Holland, J. L. (1997). *Making vocational choices* (3rd ed.). Lutz, FL: Psychological Assessment Resources, Inc.

Hough, L. M., Oswald, F. L., & Ployhart, R. E. (2001). Determinants, detection, and amelioration of adverse impact in personnel selection procedures: Issues, evidence and lessons learned. *International Journal of Selection and Assessment, 9*(1/2), 152-194.

Hughes, H. D. (1917). An interesting corn seed experiment. *Iowa Agriculturalist, 17*, 424-425.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*(1), 72-98.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. Thousand Oaks, CA: Sage.

Hursch, C. J., Hammond, K. R., & Hursch, J. L. (1964). Some methodological considerations in multiple-probability studies. *Psychological Review*, 71 (1), 42-60.

Insight. (2010). In Oxford English Dictionary. Retrieved from http://dictionary.oed.com

Jeanneret, R., & Silzer, R. (Eds.). (1998). *Individual psychological assessment: Predicting behavior in organizational setting*. San Francisco, CA: Jossey-Bass.

Jenkins, G.D., Mitra, A., Gupta, N., & Shaw, J. D. (1998). Are financial incentives related to performance?: A meta-analytic review of empirical research. *Journal of Applied Psychology, 83*(5), 777-787.

Johansson, R., & Brehmer, B. (1979). Inferences from incomplete information: A note. *Organizational Behavior and Human Performance*, 22, 141–145.

Johnson, E. J. (1988). Expertise and decision under uncertainty: Performance and process. In M.T.H. Chi, R. Glaser, & M.J. Farr (Eds.), *The nature of expertise* (pp. 209-228). Hillsdale, NJ: Erlbaum.

Johnson, J. W. (2000). A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research, 35* (1), 1-19.

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist, 64*(6), 515-526.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*(4), 237-251.

Karelaia, N. & Hogarth, R. M. (2008). Determinants of linear judgment: A meta-analysis of lens model studies. *Psychological Bulletin*, 134, 404-426 (see http://dx.doi.org/10.1037/0033-2909.134.3.404.supp for meta-analytic databases).

Kaufmann, E., & Athanasou, J. A. (2009). A meta-analysis of judgment achievement as defined by the Lens Model Equation. *Swedish Journal of Psychology*, 68(2), 99–112.

Kerce, E. W. (1985). *Boredom at work: Implications for the design of jobs with variable requirements* (Navy Personnel Research and Development Center, Technical Report No. 85-37).

Kessler, L., & Ashton, R. H. (1981). Feedback and prediction achievement in financial analysis. *Journal of Accounting Research*, 19 (1), 146-162.

Killeen, P. R. (1978). Superstition: A matter of bias, not detectability. *Science, 199*, 88-90.

Klayman, J. (1988). Learning from experience. In B. Brehmer & C.R.B. Joyce (Eds.), *Human judgment: The SJT view* (pp. 115-162). Amsterdam: North-Holland.

Klimoski, R. & Jones, R. G. (2008). Intuiting the selection context. *Industrial and Organizational Psychology, 1*, 352-354.

Kline, R. B. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research*. Washington, DC: American Psychological Association.

Koele, P. (1980). The influence of labelled stimuli on nonlinear multiple-cue probability learning. *Organizational Behavior and Human Performance*, 26, 22-31.

Komaki, J. L., Coombs, T, & Schepman, S. (1996). Motivational implications of reinforcement theory. In R. M. Steers, L. W. Porter, & G. A. Bigley (Eds.), *Motivation and leadership at work* (pp. 34-52). New York: McGraw-Hill.

Kuncel, N. R., Credé, M., & Thomas, L. L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis. *Review of Educational Research, 75*(1), 63-82.

Kuncel, N. R., Klieger, D. M., Connelly B. S., & Ones, D.S. (2010). Mechanical versus clinical data combination in selection and admissions decisions: A meta-analysis. *Journal of Applied Psychology*.

Kunreuther, H. (1969). Extensions of Bowman's theory on managerial decision-making. *Management Science*, 15, 415-439.

Kuo, B. C. (1991). *Automatic Control Systems* (6th ed.). New Jersey: Prentice Hall.

Kyllonen, P. C. & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence, 14*(4), 389-433.

Lawler, E. E. III. (1971). *Pay and organizational effectiveness*. New York: McGraw-Hill.

Lawler, E. E. III. (1973). *Motivation in the workplace*. Monterey, CA: Brooks/Cole.

Lehman, D. R., Lempert, R. O., & Nisbett, R. E. (1988). The effects of graduate training on reasoning: Formal discipline and thinking about everyday-life events. *American Psychologist, 43*, 431-442.

Lindell, M. K. (1976). Cognitive and outcome feedback in multiple-cue probability learning tasks. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(6), 739-745.

Lindell, M. K., & Stewart, T. R. (1974). The effects of redundancy in multiple-cue probability learning. *American Journal of Psychology, 87*, 393–398.

Locke, E. A., Latham, G. P., & Erez, M. (1988). The determinants of goal acceptance and commitment. *Academy of Management Review, 13*, 23-39.

Lord, F. M., & Novick, M. R. (1968). Statistical theories of mental test scores. Reading MA: Addison-Welsley Publishing Company.

Lubin, A., & Osburn, H. G. (1957). A theory of pattern analysis for the prediction of a quantitative criterion. *Psychometrika*, 22(1), 63-73.

Mandelbrot, B., & Hudson, R. L. (2006). *The (mis) behavior of markets*. New York, NY: Basic Books.

McNemar, Q. (1969). *Psychological Statistics* (4th ed.). New York, NY: John Wiley and Sons, Inc.

Meehl P. E. (1954). *Clinical versus statistical prediction*. Minneapolis, MN: University of Minnesota Press.

Meehl, P. E. (1959). A comparison of clinicians with five statistical methods of identifying psychotic MMPI profiles. *Journal of Counseling Psychology*, 6(2), 102-109.

Meehl, P. E. (1960). The cognitive activity of the clinician. *American Psychologist, 15*, 19-27.

Meehl, P. E., & Dahlstrom, W. G. (1960). Objective configural rules for discriminating psychotic from neurotic MMPI profiles. *Journal of Consulting Psychology*, 24, 575-387.

Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist, 35,* 1012–1027.

Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., & Dies, R. R., et al. (2001). Psychological testing and psychological assessment. *American Psychologist, 56*(2), 128-165.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.

Monotonic function. (2009, January 11). In Wikipedia, The Free Encyclopedia. Retrieved from http://en.wikipedia.org/wiki/Monotonic_function

Murdock, B. B., Jr. (1962). The serial position effect of free recall. *Journal of Experimental Psychology, 64,* 482-488.

Murphy, A. H. (1988). Skill scores based on mean square error and their relationships to the correlation coefficient. *Monthly Weather Review, 116*, 2417-2424.

Naylor, J. C., & Schenk, E. A. (1968). The influence of cue redundancy upon the human inference process for task of varying degrees of predictability. *Organizational Behavior and Human Performance*, 3, 47-61.

Naylor, J. C., & Shine II, L. C. (1965). A table for determining the increase in mean criterion score obtained by using a selection device. *Journal of Industrial Psychology, 3*(2), 33-42.

Norman, G.R., & Streiner, D. L. (2008). Biostatistics: The bare essentials. Shelton, CT: People's Medical Publishing House.

Nystedt, L., & Magnusson, D. (1973). Cue relevance and feedback in a clinical prediction task. *Organizational Behavior and Human Performance*, 9, 100-109.

Ogilvie, J. R., & Schmitt, N. (1979). Situational influences on linear and nonlinear use of information. *Organizational Behavior and Human Performance*, 23, 292 -306.

Ones, D. S., Viswesvaran, C., & Dilchert, S. (2004). Cognitive ability in selection decisions. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 431-468). Thousand Oaks, CA: Sage.

Phillips & Gully (2008). The role of perceptions versus reality in managers' choice of selection decision aids. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, *1*, 361-363.

Ployhart, R. E. & Holtz, B. C. (2008). The diversity–validity dilemma: strategies for reducing racioethnic and sex subgroup differences and adverse impact in selection, *Personnel Psychology, 61*(1), 153-172.

Popper, K. R., & Miller, D. W. (1983). A proof of the impossibility of inductive probability. *Nature*, 302, 687–688.

Rounds, J., Su, R., Lewis, P., & Rivkin, D. (2009). *Interest Profiler Short Form*. (U. S. Department of Labor, Employment and Training Administration, Office of Workforce Investment, Division of Workforce System Support). Washington, DC.

Ruscio, J. (2003). Holistic judgment in clinical practice: Utility or futility? [Electronic version]. *Scientific Review of Mental Health Practice, 2*(1).

Sackett, P. R., Borneman, M. J., & Connelly, B. S. (2008). High-stakes testing in higher education and employment: Appraising the evidence for validity and fairness. *American Psychologist, 63*(4), 215–227.

Sackett, P.R., & Wilk, S.L. (1994). Within-group norming and other forms of score adjustment in preemployment testing. *American Psychologist, 49*, 929-954

Sarbin, T. R., Taft, R., & Bailey, D. E. (1960). *Clinical inference and cognitive theory*. Oxford, England: Holt, Rinehart & Winston.

Sawyer, J. (1966). Measurement and prediction, clinical and statistical. *Psychological Bulletin, 66*(3), 178-200.

Schmidt, F., & Hunter, J. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262-274.

Schmitt, N., Coyle, B. W., & Saari, B. B. (1977). Types of task information feedback in multiple-cue probability learning. *Organizational Behavior and Human Performance*, 18, 316-328.

Schmitt, N., & Dudycha, A. L. (1975). The effect of cue redundancy in multiple cue probability learning. *Journal of Experimental Psychology: Human Learning and Memory, 104*, 307-315.

Shanteau, J. (1984). Some unasked questions about the psychology of expert decision makers. In M. El Haraway (Ed.), *Proceedings of the 1984 IEEE conference on systems, man, and cybernetics*. New York, NY: Institute of Electrical and Electronics Engineers, Inc.

Skitka, L. J., Mosier, K., & Burdick, M. D. (1999). Does automation bias decision-making? *International Journal of Human–Computer Studies, 51*, 991–1006.

Skitka, L. J., Mosier, K., & Burdick, M. D. (2000). Accountability and automation bias. *International Journal of Human–Computer Studies, 52*, 701–717. Skitka, L.

Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance, 6*, 649-744.

Smedslund, J. (1955). *Multiple-probability learning: An inquiry into the origins of perception*. Oslo, Norway: Oslo University Press.

Staw, B. M. (1976). Knee-deep in the big muddy: A study of escalating commitment to a chosen course of action. *Organizational Behavior and Human Performance, 16*(1), 27-44.

Steinmann, D. (1976). The effects of cognitive feedback and task complexity in multiple cue probability learning. *Organizational Behavior & Human Performance, 15*, 168-179.

Stewart, T. R. (1976). Components of correlation and extensions of the lens model equation. *Psychometrika, 41*, 101-120.

Stewart, T. R. (1988). Judgment analysis: Procedures. In B. Brehmer & C.R.B. Joyce (Eds.), *Human judgment: The SJT view* (pp. 41-74). Amsterdam: North-Holland.

Stewart, T. R. (1990). A decomposition of the correlation coefficient and its use in analyzing forecasting skill. *Weather and Forecasting, 5,* 661-666.

Stewart, T. R. (1994, November). *What does G measure? What should be done about it?* Presented at the 10th Annual International Invitational Meeting of the Brunswik Society, St. Louis, MO.

Stewart, T. R., & Ely, D. W. (1984). *Range sensitivity: A necessary condition and a test for the validity of weights* (Report No. 270). Center for Research n Judgment and Policy, University of Colorado, Boulder, CO.

Steyn, M. (2003, December 9). Rummy speaks the truth, not gobbledygook. *Daily Telegraph*. Retrieved from http://www.telegraph.co.uk/comment/personal-view/3599959/Rummy-speaks-the-truth-not-gobbledygook.html.

Taylor, H. C., & Russell, J. T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection: Discussion and tables. *Journal of Applied Psychology, 23*, 565-578.

Tellegen A., Grove W. M., & Waller N. G. (1987). *Inventory of personal characteristics #7 (IPC-7).* Unpublished materials, University of Minnesota.

Thoreau, H. D. (1882). *Walden*. Cambridge, MA: Riverside Press.

Thorndike, E. L. (1918). Fundamental theorems in judging men. *Journal of Applied Psychology, 2*, 67-76.

Tucker, L. R. (1964). A suggested alternative formulation in the developments by Hursch, Hammond and Hursch and by Hammond, Hursch and Todd. *Psychological Review*, 71, 528-530.

Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124-1131.

Tversky, A. & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*, 453-458.

Vonnegut, Jr., K. (1985). *Galápagos*. New York: Dell Publishing.

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin, 117*(2), 250-270.

Vroom, V. H. (1964). *Work and motivation*. New York: Wiley.

Wiggins, N., & Kohen, E. S. (1971). Man versus model of man revisited: The forecasting of graduate school success. *Journal of Personality and Social Psychology*, 19 (1), 100-106.

Willings, D. (1984). Considerations in counseling the creatively gifted. *International Journal for the Advancement of Counseling*, 7, 3-13.

Wright, P. (1974). The harassed decision maker: Time pressures, distractions, and the use of information. *Journal of Applied Psychology*, 59, 555-561.

Yntema, D. B., & Torgerson, W. S. (1961). Man-computer cooperation in decisions requiring common sense. *IRE Transactions of the Professional Group on Human Factors in Electronics, HFE-2,* 20-26.

Zedeck, S., Tziner, A., & Middlestadt S. E. (1983). Individual validity and reliability: An individual analysis approach. *Personnel Psychology, 36*, 355-370.

APPENDICES

Appendix A

Differences Between Designs of Fall 2009 and Spring 2010 Studies

	Aspect of Studies	Difference Between Design of Fall 2009 Study and Design of Spring 2010 Study	Notes
1	Number of subjects		
2	Number of blocks of predictions made by subjects	4 blocks are analyzed for Fall 2009, whereas 5 blocks are analyzed for Spring 2010. For Fall 2009, the 5th block was eliminated from analyses for reasons discussed elsewhere.	
3	Item ordering for the prediction task (to minimize ordering effects)	Fall 2009 employed only 2 different item orderings, whereas Spring 2010 employed a fully randomized blocks design (5! permutations for each of 3 experimental conditions = 120 orderings per experimental condition).	
4	Description of the jobs for which hypothetical applicants were applying	Fall 2009 indicates that the job can vary across applicants, whereas Spring 2010 indicates that the job is the same across applicants.	Compare Appendix B (Fall 2009 Instructions) to Appendix F (Spring 2010 Instructions)
5	Simulated Cues (Cognitive Ability Test Score & Interesting/Boring Nature of Job)	Same, except Fall 2009 used a 0-6 anchored Likert scale to measure the cues, whereas Spring 2010 used a percentile scale for measuring cognitive ability test score and a dichotomous scale for measuring how interesting/boring the job would be.	Compare Appendix B (Fall 2009 Instructions) to Appendix F (Spring 2010 Instructions); or compare Appendix C (Fall 2009 Sample of Prediction Profiles Form) to Appendix G (Spring 2010 Sample of Prediction Profiles Form)
6	Simulated Criterion (Job Performance)	Same, except Fall 2009 used a 0-6 anchored Likert scale to measure the criterion, whereas Spring 2010 used a percentile scale.	Compare Appendix B (Fall 2009 Instructions) to Appendix F (Spring 2010 Instructions); or compare Appendix C (Fall 2009 Sample of Prediction Profiles Form) to Appendix G (Spring 2010 Sample of Prediction Profiles Form)

Aspect of Studies		Difference Between Design of Fall 2009 Study and Design of Spring 2010 Study	Notes
7	Disordinal Interaction	Same, except the disordinal interaction is generally stronger for Spring 2010.	See Tables 3 and 4 for data comparison.
8	Feedback Conditions	For Fall 2009, across all 3 experimental conditions, the graphical feedback for the first 4 blocks was based on blocks of profiles selectively chosen so that for each block the statistical relationships among cues and between cues and the job performance criterion at least vaguely resembled the statistical relationships for the population of data (including the graphical appearance of disordinal interaction). All blocks for Spring 2010 were randomly chosen.	See functional forms in Appendix E (Fall 2009 Task Information Feedback).
		For Fall 2009, across all 3 experimental conditions, there appear scatterplots (with best-fitting lines) that illustrate (a) cue-criterion relationships ignoring the effects of the other cue and (b) cue intercorrelations. No such displays appear for Spring 2010.	Compare Appendix E (Fall 2009 Task Information Feedback) to Appendix I (Spring 2010 Task Information Feedback).
		For Fall 2009, across all 3 experimental conditions, graphical feedback appeared partially in color. For Spring 2010, all graphical feedback appeared in grayscale.	
		For Spring 2010, across all 3 experimental treatments, and for the data underlying each block (i.e., for each sample of profiles), subjects always received as graphical feedback scatterplots showing the relationship of cognitive ability to job performance for interesting jobs and then separately for boring jobs. For Fall 2009, only in conditions 2 and 3 was the disordinal interaction graphically depicted; it appeared for both boring and interesting jobs as 2 differently colored lines intersecting in the same line plot (with no scatter shown).	Compare Appendix I (Spring 2010 Task Information Feedback) to Appendix E (Fall 2009 Task Information Feedback).
		For Spring 2010, for feedback conditions 2 and 3 only, and for each block, subjects received detailed instructions with a detailed example of how to interpret a scatterplot of a disordinal interaction. For Fall 2009, instructions and the example for the disordinal interaction were provided for feedback conditions 2 and 3 only, but such instructions and feedback were not as detailed as in Spring 2010.	Compare Appendix I (Spring 2010 Task Information Feedback) to Appendix E (Fall 2009 Task Information Feedback).
		For Spring 2010, for the third feedback condition, subjects were told that they should use the graphical feedback depicting the disordinal interaction to make predictions; for Spring 2010, for the second feedback condition, subjects were told that the graphical feedback depicting the disordinal interaction might be important to making predictions. For Fall 2009, no such prescriptions or warnings were provided to subjects in any feedback condition.	Compare Appendix I (Spring 2010 Task Information Feedback) to Appendix E (Fall 2009 Task Information Feedback).
9	Response Scale for RIASEC Interest Profiler Form	4-point scale for Fall 2009 and 5-point scale for Spring 2010.	See Interest Profiler Form in Appendix O.

	Aspect of Studies	Difference Between Design of Fall 2009 Study and Design of Spring 2010 Study	Notes
10	Dependent Variables (DVs) Analyzed	Same DVs were analyzed, except longitudinal DVs for Spring 2010 include a value for a fifth time point (i.e., fifth block).	See Table 5 for a complete list of analyzed variables and their descriptions.
11	Independent Variables (IVs) Analyzed, Including Individual Differences Variables	Same IVs were analyzed, except longitudinal IVs for Spring 2010 include a value for a fifth time point (i.e., fifth block).	See Table 5 for a complete list of analyzed variables and their descriptions.

Appendix B

Instructions to Judgment Task (Fall 2009)

Prize:

The participant in this study who, on average, predicts job performance of the job applicants better than every other participant does will win $75 cash. The participant who performs second-best will win $50 cash, and the participant who performs third-best will win $25 cash. Participants who win will be contacted when this study is completed, so please be sure that we have contact information that will still be valid in August 2010.

Introduction:

Imagine that you are a hiring manager responsible for predicting job performance of 125 job applicants.

Cognitive Ability Test Score:

For each job applicant, you will be told how that applicant scored on a recent cognitive ability (intelligence) test. All applicants took the same exact test. Assume that higher cognitive ability is associated with higher job performance. Please note that there exists empirical evidence that demonstrates a moderate to strong relationship between cognitive ability and job performance (e.g., Schmidt & Hunter, 1998). An applicant's cognitive ability test results are scored from 0 to 6. A higher score indicates a higher level of cognitive ability. A lower score indicates a lower level of cognitive ability. The scoring system looks like the following:

|------------------|-------------------|------------------|--------------------|-------------------|------------------|

0	1	2	**3**	4	5	**6**
Extremely low cognitive ability	Moderately below average cognitive ability	Slightly below average cognitive ability	Average cognitive ability	Slightly above average cognitive ability	Moderately above average cognitive ability	Extremely high cognitive ability

How Intellectually Interesting or Boring a Job Is:

Job applicants are not necessarily applying for the same exact job. From the perspective of a job expert, the jobs for which people are applying will differ from each other in terms of how intellectually interesting or boring the jobs are. In other words, please assume that most reasonable people familiar with the job would agree about how intellectually interesting or boring a job is. Assume that a more interesting (and less boring) job is associated with higher job performance. Also assume that employees with higher cognitive ability tend to be in more interesting (and less boring) jobs. The job expert scored from 0 to 6 how intellectually interesting or boring that job is. A higher score indicates that a job is more intellectually interesting or less intellectually boring. A lower score indicates that a job is less intellectually interesting or more intellectually boring. The scoring system looks like the following:

|------------------|-------------------|------------------|--------------------|-------------------|------------------|

0	1	2	**3**	4	5	**6**
Extremely intellectually boring	Moderately intellectually boring	Slightly intellectually boring	Average level intellectually interesting/ boring	Slightly intellectually interesting	Moderately intellectually interesting	Extremely intellectually interesting

Your Task:

Your task is to predict how each job applicant would perform in the job if hired. Know that perfect prediction is extremely unlikely. Just do your best. You are to score each applicant from 0 to 6. A higher score indicates that a job applicant would perform better in the job. A lower score indicates that a job applicant would perform worse in the job. The scoring system looks like the following:

|------------------|-------------------|------------------|--------------------|-------------------|------------------|

0	1	2	**3**	4	5	**6**
You expect that the job applicant would perform extremely poorly in the job if hired	You expect that the job applicant would perform moderately poorly in the job if hired	You expect that the job applicant would perform slightly poorly in the job if hired	You expect that the job applicant would perform average in the job if hired	You expect that the job applicant would perform slightly well in the job if hired	You expect that the job applicant would perform moderately well in the job if hired	You expect that the job applicant would perform extremely well in the job if hired

For each job applicant, please indicate your response clearly and neatly on the scantron form.

You have received "profiles" for the 125 job applicants. There is 1 profile for each job applicant. A profile is simply a combination of cognitive ability score and how intellectually interesting or boring the job is.

Feedback:

After every set of 25 profiles (except for the very last set), you will receive the feedback information briefly described below regarding the task that you are doing. This feedback information might not appear in the exact order below. Feedback information will appear as illustrations (i.e., pictures), sometimes followed by an explanation to help you interpret the illustrations. The feedback information that you are given describes the immediately previous 25 judgments that you were asked to make.

1. Predictability: This information tells you how well pre-existing mathematical formulas would have predicted job performance.

2. Redundancy: This information tells you the extent to which (a) cognitive ability test score and (b) how intellectually interesting or boring the job is tell you the same exact thing about how well the job applicant will perform in the job. In making a prediction, a hiring manager sometimes will use pieces of information that are redundant to some extent. However, the more that those pieces of information independently tell you something about an outcome (e.g., job performance), the more useful those pieces of information can be.

3. Relative Importance: This information tells you which of the following is more important to predicting job performance: (a) cognitive ability test score, (b) how intellectually interesting or boring the job is, or (c) some other information. Assuming that one of these three things will be more important than the others, this information also will tell you by how much one of these things is more important than the others.

4. Functional Form: This information will tell you something about how (a) cognitive ability test score relates to job performance, (b) how intellectually interesting or boring the job is relates to job performance, and (c) how cognitive ability test score relates to how intellectually interesting or boring the job is.

If you have any questions, please ask the person leading the study.

<u>Appendix C</u>

<u>Sample of Prediction Profiles Form (Fall 2009)</u>

Below are the cognitive ability test scores for each of 25 job applicants.
For each applicant it is noted below how intellectually interesting / boring the job is for which the applicant is applying.
On the scantron form, please indicate your prediction of each applicant's job performance if the applicant were hired.
Below at right are the rating scales that you should use for job performance, cognitive ability, and how intellectually interesting / boring the job is.
Each response on the scantron form should correspond to the job applicant's identification number provided below.
Please do not leave this page until you have made predictions for all 25 job applicants.

Job applicant's identification number:	Job applicant's <u>cognitive ability</u> test score (from 0 to 6):	How <u>intellectually interesting / boring</u> the job is that the applicant is applying for (from 0 to 6):	Your prediction of the applicant's <u>job performance</u> if the applicant were hired? (from 0 to 6):
1	5	5	?
2	2	5	?
3	2	1	?
4	3	1	?
5	1	1	?
6	4	1	?
7	3	1	?
8	3	5	?
9	4	2	?
10	3	1	?
11	4	4	?
12	2	1	?
13	5	1	?
14	3	4	?
15	3	1	?
16	5	4	?
17	1	4	?
18	3	6	?
19	1	1	?
20	3	2	?
21	1	4	?
22	4	1	?
23	3	5	?
24	1	4	?
25	3	5	?

<u>Job Performance Rating Scale:</u>

0	1	2	3	4	5	6
You expect that the job applicant would perform <u>extremely poorly</u> in the job if hired	You expect that the job applicant would perform <u>moderately poorly</u> in the job if hired	You expect that the job applicant would perform <u>slightly poorly</u> in the job if hired	You expect that the job applicant would perform <u>average</u> in the job if hired	You expect that the job applicant would perform <u>slightly well</u> in the job if hired	You expect that the job applicant would perform <u>moderately well</u> in the job if hired	You expect that the job applicant would perform <u>extremely well</u> in the job if hired

<u>Cognitive Ability Rating Scale:</u>

0	1	2	3	4	5	6
Extremely low cognitive ability	Moderately below average cognitive ability	Slightly below average cognitive ability	Average cognitive ability	Slightly above average cognitive ability	Moderately above average cognitive ability	Extremely high cognitive ability

<u>Rating Scale for How Intellectually Interesting / Boring the Job Is that the Applicant Is Applying For:</u>

0	1	2	3	4	5	6
Extremely intellectually boring	Moderately intellectually boring	Slightly intellectually boring	Average level of being intellectually interesting/ boring	Slightly intellectually interesting	Moderately intellectually interesting	Extremely intellectually interesting

Appendix D

Major Differences Among the Feedback Treatment Conditions (Fall 2009) (See Appendix E for the Graphical Feedback Itself)

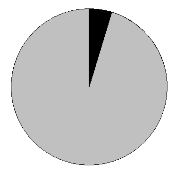
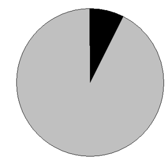
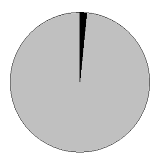
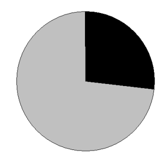
	Condition 1	**Condition 2**	**Condition 3**
1. Purpose	To determine the extent to which subjects can identify and effectively utilize the disordinal interaction without being explicitly told it might exist or does exist.	To determine the extent to which subjects can identify and effectively utilize the disordinal interaction if explicitly told that it does exist.	Same as in Condition 2, but explicit feedback about the disordinal interaction is received 1 block sooner.
2. Feedback: Scatterplots and Line Graphs	Scatterplots of: 1. Relationship between cognitive ability and job performance; 2. Relationship between how interesting/boring the job is and job performance; and 3. Cognitive ability and how interesting/boring the job is. Scatterplots include best-fitting lines. Detailed instructions with example in color accompany the scatterplots to aid interpretation.	Same scatterplots and scatterplot instructions with example as in Condition 1. Line graph explicitly showing the disordinal interaction (Y = actual job performance; X = cognitive ability test score). There are 2 separate and differently colored lines: one for when the applicant is expected to find the job interesting and one for when the applicant is expected to find the job boring. Depicts data for the immediately prior block. Line graph feedback is received starting after block 2. Detailed example of a disordinal interaction for an unrelated context. Existence of disordinal interaction in the prediction data is mentioned as a possibility only.	Same scatterplots and scatterplot instructions with example as in Condition 1. Same line graphs as in Condition 2, but received 1 block sooner (i.e., after block 1). Same detailed example of disordinal interaction as in Condition 2. Same portrayal as Condition 2 of the possible existence of a disordinal interaction in the prediction data.

	Condition 1	Condition 2	Condition 3
3. Feedback: Pie Charts	Pie charts: Provide task information feedback without the disordinal interaction. Feedback includes: 1. How predictable job performance is when using a linear mechanical equation <u>(with nothing about predictability for the interaction mechanical equation)</u> 2. C_{xj} relative weights (dominance analysis) for (1) cognitive ability test score and (2) whether the applicant is expected to find the job interesting or boring <u>(with nothing about the interaction term)</u> 3. Redundancy (inter-correlation) for (1) cognitive ability test score and (2) whether the applicant is expected to find the job interesting or boring.	Pie charts: Same as in Condition 1 until block 2. After block 2, designed to emphasize importance of the disordinal interaction. Feedback includes: 1. How predictable job performance is when using a linear mechanical equation <u>versus how predictable job performance is when using an interaction mechanical equation</u> 2. C_{xj} relative weights (dominance analysis) for (1) cognitive ability test score, (2) whether the applicant is expected to find the job interesting or boring, <u>and (3) the interaction term</u>. 3. Redundancy (inter-correlation) information same as in Condition 1.	Same as in Condition 2 with Condition 2's block 2+ design starting with block 1. (Disordinal interaction is emphasized with initial feedback.)

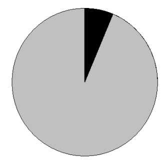
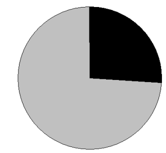
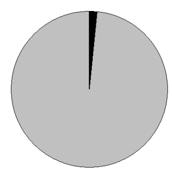
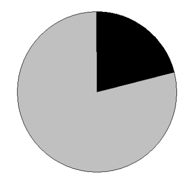
Appendix E

Task Information Feedback (Fall 2009)

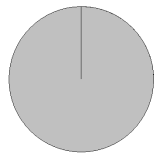
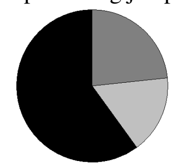
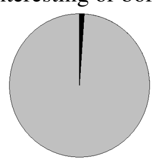
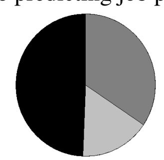
(See Appendix D for a Narrative Description of Major Differences Among the Feedback Treatment Conditions)

	PREDICTABILITY (R_e^2) **(Excluding the Interaction):**	**PREDICTABILITY (R_e^2)** **(Including the Interaction):**
After the first [fourth] 25 judgments:	How predictable was job performance?  Total pie = perfect prediction (in an ideal world) Black slice = the actual predictability when using an equation created in advance The bigger the black slice, the more predictable that job performance is.	How predictable was job performance?  Total pie = perfect prediction (in an ideal world) Black slice = the actual predictability when using an equation created in advance The bigger the black slice, the more predictable that job performance is.
After the second [third] 25 judgments:	How predictable was job performance?  Total pie = perfect prediction (in an ideal world) Black slice = the actual predictability when using an equation created in advance The bigger the black slice, the more predictable that job performance is.	How predictable was job performance?  Total pie = perfect prediction (in an ideal world) Black slice = the actual predictability when using an equation created in advance The bigger the black slice, the more predictable that job performance is.

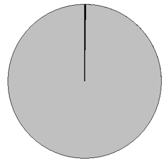
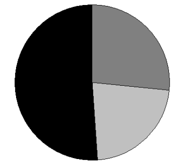
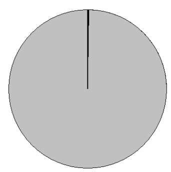
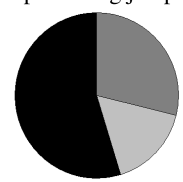
Notes. The "[" and "]" brackets indicate a reverse-ordering of blocks (i.e., sets of 25 trials or judgments) and judgments within blocks that will be used to help reduce and/or explain the impact of ordering effects on study results. For all of Condition 1 and the first half (first 50 predictions) of Condition 2, the subjects will be given only the predictability feedback that includes the interaction. This information is given in lieu of predictability feedback that excludes the interaction, because predictability feedback that includes the interaction establishes a higher goal for subjects to achieve. Recall that the goal is for people to outperform the simple linear model. For the second half of Condition 2 (second 50 predictions) and all of Condition 3, the subjects are given side-by-side predictability feedback (the pie chart) that includes the interaction and predictability feedback (the pie chart) that excludes the interaction, with the predictability feedback types specifically labeled as such. Providing both pieces of information for subjects to compare is intended to encourage subjects to employ the interaction. For the second half of Condition 2 and all of Condition 3, subjects already have been taught about the disordinal interaction before receiving predictability feedback, because initial feedback will consist of a plot of the disordinal interaction with instructions for interpreting it. (The functional forms feedback for the disordinal interactions, with the instructions to subjects about how to interpret that feedback, is described below.)

	PREDICTABILITY (R_e^2) **(Excluding the Interaction):**	**PREDICTABILITY (R_e^2)** **(Including the Interaction):**
After the third [second] 25 judgments:	How predictable was job performance?  Total pie = perfect prediction (in an ideal world) Black slice = the actual predictability when using an equation created in advance The bigger the black slice, the more predictable that job performance is.	How predictable was job performance?  Total pie = perfect prediction (in an ideal world) Black slice = the actual predictability when using an equation created in advance The bigger the black slice, the more predictable that job performance is.
After the fourth [first] 25 judgments:	How predictable was job performance?  Total pie = perfect prediction (in an ideal world) Black slice = the actual predictability when using an equation created in advance The bigger the black slice, the more predictable that job performance is.	How predictable was job performance?  Total pie = perfect prediction (in an ideal world) Black slice = the actual predictability when using an equation created in advance The bigger the black slice, the more predictable that job performance is.

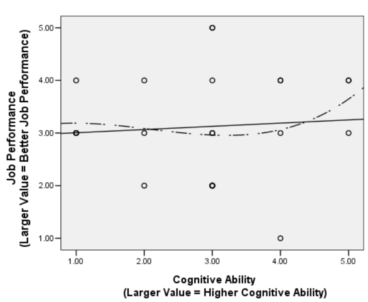
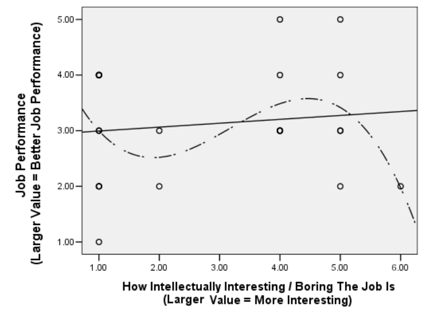
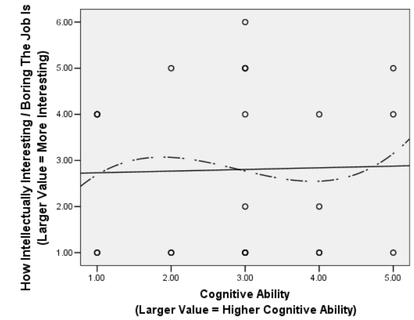
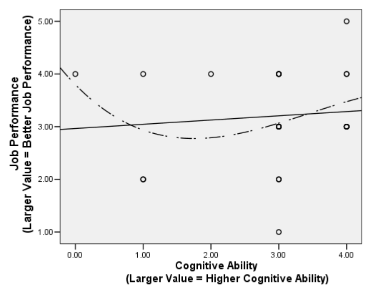
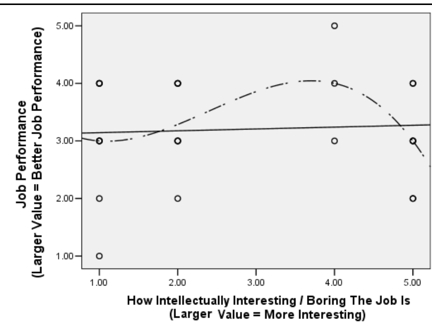
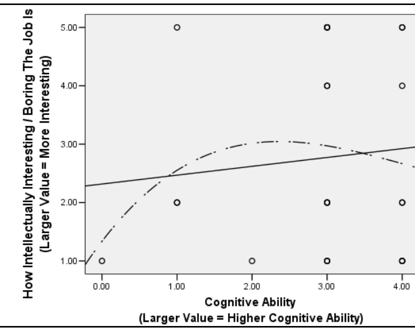
Notes. The "[" and "]" brackets indicate a reverse-ordering of blocks (i.e., sets of 25 trials or judgments) and judgments within blocks that were used to help reduce and/or explain the impact of ordering effects on study results. For all of Condition 1 and the first half (first 50 predictions) of Condition 2, the subjects were given only the predictability feedback that included the interaction. This information was given in lieu of predictability feedback that excluded the interaction, because predictability feedback that includes the interaction established a higher goal for subjects to achieve. Recall that the goal is for people to outperform the simple linear model. For the second half of Condition 2 (second 50 predictions) and all of Condition 3, the subjects were given side-by-side predictability feedback (the pie chart) that included the interaction and predictability feedback (the pie chart) that excluded the interaction, with the predictability feedback types specifically labeled as such. Providing both pieces of information for subjects to compare was intended to encourage subjects to employ the interaction. For the second half of Condition 2 and all of Condition 3, subjects were already taught about the disordinal interaction before receiving predictability feedback, because initial feedback consisted of a plot of the disordinal interaction with instructions for interpreting it. (The functional forms feedback for the disordinal interactions, with the instructions to subjects about how to interpret that feedback, is described below.)

	REDUNDANCY ($r^2_{x1, x2}$):	**RELATIVE IMPORTANCE (C_{xj}):**
After the first [fourth] 25 judgments:	How redundant was (1) the cognitive ability test score with (2) how cognitively interesting or boring the job is?  Total pie = total redundancy (i.e., if these two pieces of information told you the same exact thing about job performance) Black slice = the actual extent of redundancy The bigger the black slice, the more that these two pieces of information tell you the same exact thing about job performance.	How important were (1) the cognitive ability test score and (2) how cognitively interesting or boring the job is to predicting job performance?  Total pie = everything that should be considered when predicting job performance Dark grey slice = how cognitively interesting or boring the job is Light gray slice = cognitive ability test score Black slice = other information {the disordinal interaction}←COND. 3 ONLY The bigger the slice, the more important that it is to predicting job performance.
After the second [third] 25 judgments:	How redundant was (1) the cognitive ability test score with (2) how cognitively interesting or boring the job is?  Total pie = total redundancy (i.e., if these two pieces of information told you the same exact thing about job performance) Black slice = the actual extent of redundancy The bigger the black slice, the more that these two pieces of information tell you the same exact thing about job performance.	How important were (1) the cognitive ability test score and (2) how cognitively interesting or boring the job is to predicting job performance?  Total pie = everything that should be considered when predicting job performance Dark grey slice = how cognitively interesting or boring the job is Light gray slice = cognitive ability test score Black slice = other information {the disordinal interaction}←COND. 3 ONLY The bigger the slice, the more important that it is to predicting job performance.

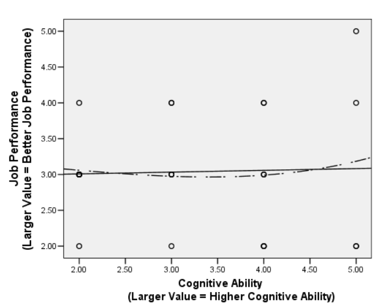
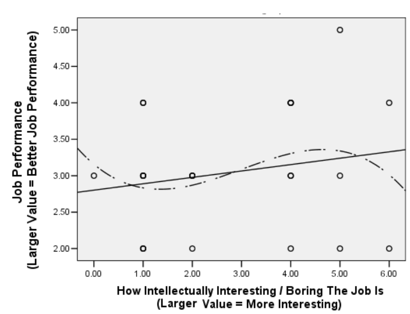
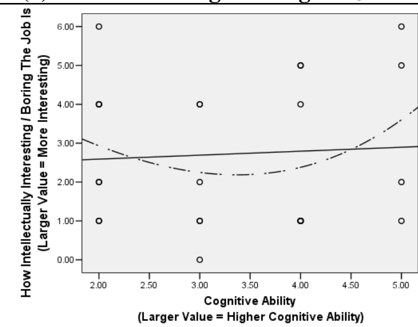
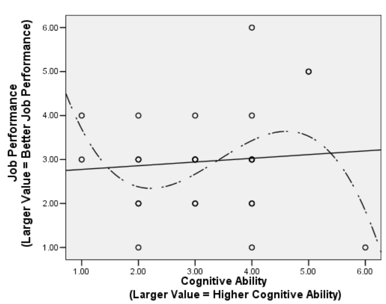
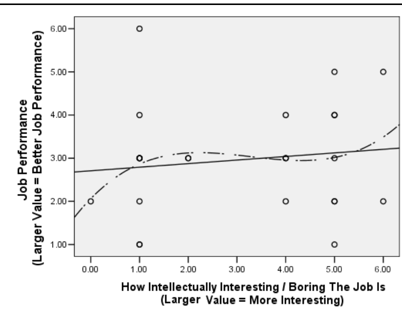
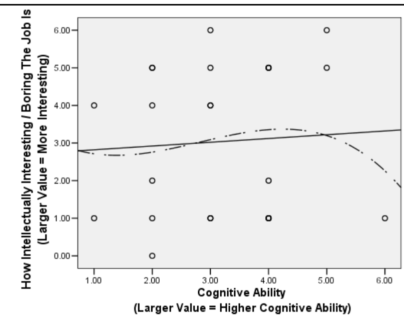
Notes. The "[" and "]" brackets indicate a reverse-ordering of blocks (i.e., sets of 25 trials or judgments) and judgments within blocks that were used to help reduce and/or explain the impact of ordering effects on study results. For relative importance, the black slice = "other information" for all of Condition 1 and the first half (first 50 predictions) of Condition 2. For the second half of Condition 2 (second 50 predictions) and all of Condition 3, the black slice = "the interaction in which the relationship of cognitive ability to job performance depends upon how interesting/boring the job is". The "{" and "}" brackets indicate information that was provided to subjects only for the second half of Condition 2 and all of Condition 3, when subjects were encouraged to learn and employ the disordinal interaction. For the second half of Condition 2 and all of Condition 3, subjects already had been taught about the disordinal interaction before receiving redundancy and relative importance feedback, because initial feedback consisted of a plot of the disordinal interaction with instructions for interpreting it. (The functional forms feedback for the disordinal interactions, with the instructions to subjects about how to interpret that feedback, is described below.)

	REDUNDANCY ($r^2_{x1, x2}$):	**RELATIVE IMPORTANCE (C_{xj}):**
After the third [second] 25 judgments:	How redundant was (1) the cognitive ability test score with (2) how cognitively interesting or boring the job is?  Total pie = total redundancy (i.e., if these two pieces of information told you the same exact thing about job performance) Black slice = the actual extent of redundancy The bigger the black slice, the more that these two pieces of information tell you the same exact thing about job performance.	How important were (1) the cognitive ability test score and (2) how cognitively interesting or boring the job is to predicting job performance?  Total pie = everything that should be considered when predicting job performance Dark grey slice = how cognitively interesting or boring the job is Light gray slice = cognitive ability test score Black slice = other information {the disordinal interaction}←COND. 3 ONLY The bigger the slice, the more important that it is to predicting job performance.
After the fourth [first] 25 judgments:	How redundant was (1) the cognitive ability test score with (2) how cognitively interesting or boring the job is?  Total pie = total redundancy (i.e., if these two pieces of information told you the same exact thing about job performance) Black slice = the actual extent of redundancy The bigger the black slice, the more that these two pieces of information tell you the same exact thing about job performance.	How important were (1) the cognitive ability test score and (2) how cognitively interesting or boring the job is to predicting job performance?  Total pie = everything that should be considered when predicting job performance Dark grey slice = how cognitively interesting or boring the job is Light gray slice = cognitive ability test score Black slice = other information {the disordinal interaction}←COND. 3 ONLY The bigger the slice, the more important that it is to predicting job performance.

Notes. Notes. The "[" and "]" brackets indicate a reverse-ordering of blocks (i.e., sets of 25 trials or judgments) and judgments within blocks that were used to help reduce and/or explain the impact of ordering effects on study results. For relative importance, the black slice = "other information" for all of Condition 1 and the first half (first 50 predictions) of Condition 2. For the second half of Condition 2 (second 50 predictions) and all of Condition 3, the black slice = "the interaction in which the relationship of cognitive ability to job performance depends upon how interesting/boring the job is". The "{" and "}" brackets indicate information that was provided to subjects only for the second half of Condition 2 and all of Condition 3, when subjects were encouraged to learn and employ the disordinal interaction. For the second half of Condition 2 and all of Condition 3, subjects already had been taught about the disordinal interaction before receiving redundancy and relative importance feedback, because initial feedback consisted of a plot of the disordinal interaction with instructions for interpreting it. (The functional forms feedback for the disordinal interactions, with the instructions to subjects about how to interpret that feedback, is described below.)

<table>
<tr><td></td><td colspan="3">FUNCTIONAL FORMS:</td></tr>
<tr><td></td><td>Relationship Between
(1) Cognitive Ability and
(2) Job Performance</td><td>Relationship Between
(1) How Interesting / Boring the Job Is and
(2) Job Performance</td><td>Relationship Between
(1) Cognitive Ability and
(2) How Interesting / Boring the Job Is</td></tr>
<tr><td>After the first [fourth] 25 judgments:</td><td>
</td><td>
</td><td>
</td></tr>
<tr><td>After the second [third] 25 judgments:</td><td>
</td><td>
</td><td>
</td></tr>
</table>

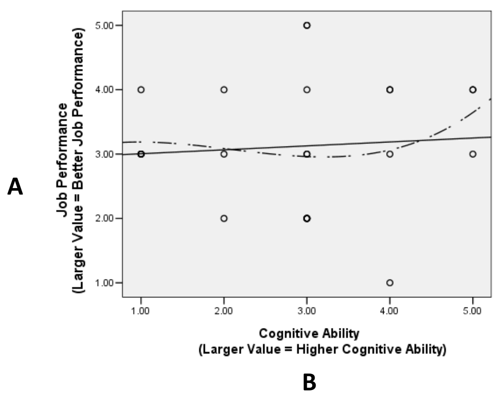
<u>Notes</u>. The "[" and "]" brackets indicate a reverse-ordering of blocks (i.e., sets of 25 trials or judgments) and judgments within blocks that were used to help reduce and/or explain the impact of ordering effects on study results. The dashed, curved lines are cubic models fitted to the data.

<table>
<tr><td></td><td colspan="3">FUNCTIONAL FORMS:</td></tr>
<tr><td></td><td>Relationship Between
(1) Cognitive Ability and
(2) Job Performance</td><td>Relationship Between
(1) How Interesting / Boring the Job Is and
(2) Job Performance</td><td>Relationship Between
(1) Cognitive Ability and
(2) How Interesting / Boring the Job Is</td></tr>
<tr><td>After the third [second] 25 judgments:</td><td>
</td><td>
</td><td>
</td></tr>
<tr><td>After the fourth [first] 25 judgments:</td><td>
</td><td>
</td><td>
</td></tr>
</table>

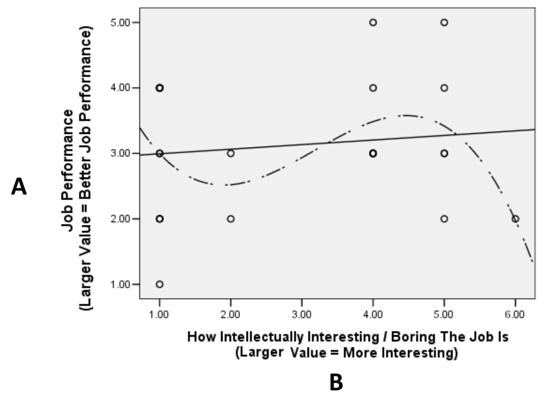
<u>Notes</u>. The “[“ and “]” brackets indicate a reverse-ordering of blocks (i.e., sets of 25 trials or judgments) and judgments within blocks that were used to help reduce and/or explain the impact of ordering effects on study results. The dashed, curved lines are cubic models fitted to the data.

EXAMPLE OF INSTRUCTIONS TO SUBJECTS ABOUT HOW THEY SHOULD INTERPRET THE SCATTERPLOTS:

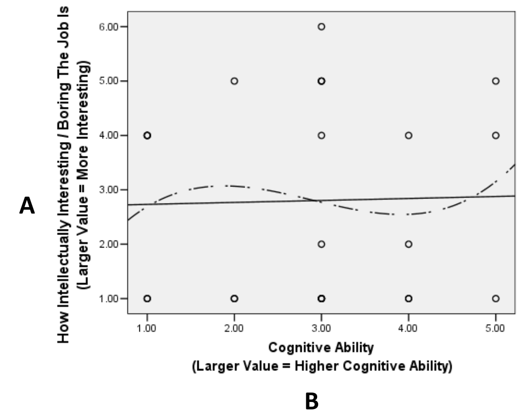
Relationship Between
(1) Cognitive Ability and (2) Job Performance



Relationship Between
(1) How Interesting / Boring the Job Is and
(2) Job Performance



Relationship Between
(1) Cognitive Ability and
(2) How Interesting / Boring the Job Is

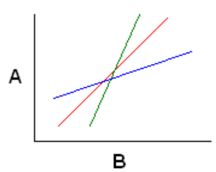


How to interpret the 3 illustrations above:

<table>
<tr><td colspan="3">Each dot (each tiny circle) above represents the scores for a single job applicant. Each line (straight and curved) is an attempt to fit (to summarize) the pattern of dots and thereby represent (summarize) each relationship.</td></tr>
<tr>
<td>If A increases in value as B increases in value, then the straight line in the illustration will tend to look similar to one of the 3 colored lines below.

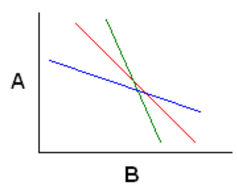
The more tilted (slanted) the straight line, the stronger the relationship is between A and B.

The red line below represents maximum tilt.

</td>
<td>If A increases in value as B decreases in value, then the straight line in the illustration will tend to look similar to one of these 3 colored lines below.

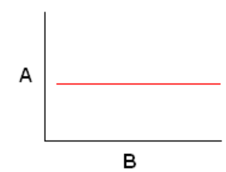
The more tilted (slanted) the straight line, the stronger the relationship is between A and B.

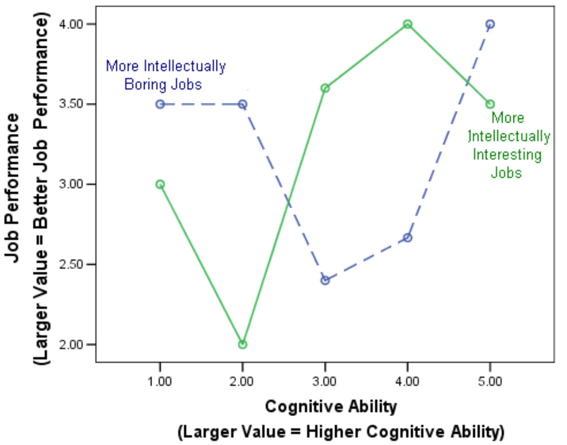
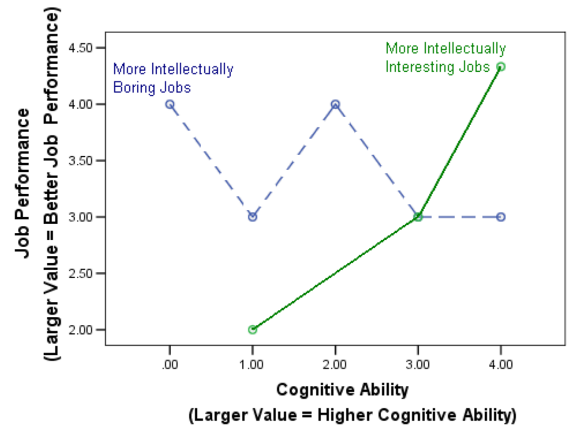
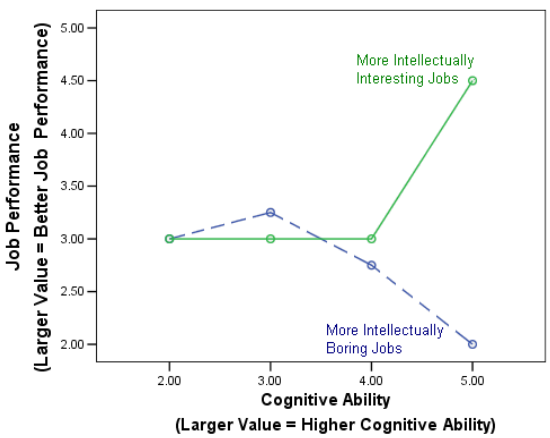
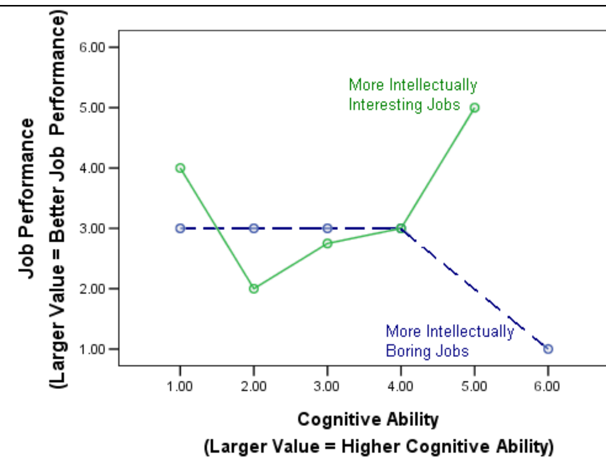
The red line below represents maximum tilt.

</td>
<td>The flatter (the less tilted or slanted) the straight line is, the weaker the relationship is between A and B.

If A and B are completely unrelated to each other, then the straight line in the illustration will be perfectly horizontal (flat) like the red line immediately below.

So the red line below represents a lack of tilt.

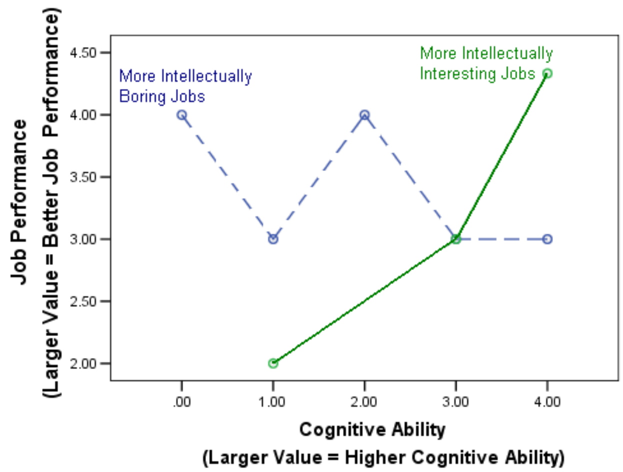
</td>
</tr>
<tr><td colspan="3">A curved, dashed line (see above) represents another way to relate A to B. A straight line usually provides helpful information even when it does not fit the dots as well as a curved line does. However, it is up to you to decide whether and by how much a curved line or some other alternative to a straight line represents the relationship of A to B better than a straight line does. For example, if a curved line represents the relationship of A to B far better than the straight line does, you might choose to completely ignore the straight line. It also is up to you what, if anything, to do with the information provided by the straight and curved lines in making your predictions of job performance.</td></tr>
</table>

<table>
<tr><th colspan="4">FUNCTIONAL FORMS:</th></tr>
<tr><td></td><td>Disordinal Interaction: Relationship Between Cognitive Ability and Job Performance When Considering How Interesting / Boring the Jobs Are</td><td></td><td>Disordinal Interaction: Relationship Between Cognitive Ability and Job Performance When Considering How Interesting / Boring the Jobs Are</td></tr>
<tr><td>After the first [fourth] 25 judgments:</td><td>
</td><td>After the second [third] 25 judgments:</td><td>
</td></tr>
<tr><td>After the third [second] 25 judgments:</td><td>
</td><td>After the fourth [first] 25 judgments:</td><td>
</td></tr>
</table>

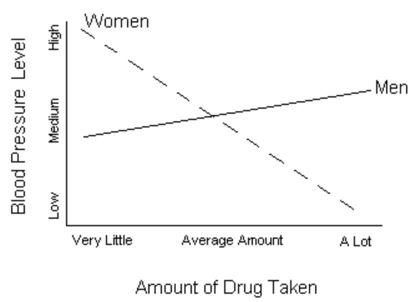
<u>Notes</u>. None of these displays were provided in Condition 1. In Condition 2, these displays were provided only after the second, third, and fourth blocks (after the first 50 predictions are made). In Condition 3, these displays were provided after every block (after every 25 predictions) except for the last block. The "[" and "]" brackets above indicate a reverse-ordering of blocks (i.e., sets of 25 trials or judgments) and judgments within blocks that were used to help reduce and/or explain the impact of ordering effects on study results. Plotted Job Performance scores were actually estimated marginal means produced by SPSS. One line segment for 2 of the trends was produced via interpolation.

EXAMPLE OF INSTRUCTIONS TO SUBJECTS ABOUT HOW THEY SHOULD INTERPRET THE DISORDINAL INTERACTION FUNCTIONAL FORM PLOTS:

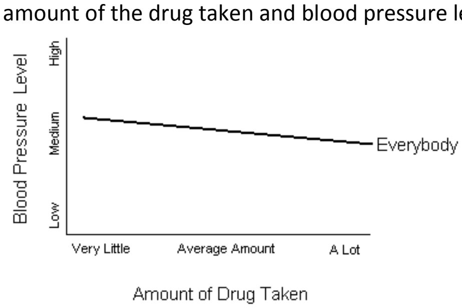
Disordinal Interaction: Relationship Between (1) Cognitive Ability and (2) Job Performance
When Considering How Interesting / Boring the Jobs Are



How to interpret the illustration above:
Like in other situations where predictions are made, it is possible that what one piece of information tells you about an outcome depends upon what another piece of information tells you about that outcome. One example of an interaction (specifically a "disordinal" interaction) is depicted below. It shows that as women take more of a drug to keep blood pressure levels low, their blood pressure levels drop. However, as men take more of the same drug, their blood pressure levels increase.



What if we had ignored gender and just looked at everybody together? We would have seen that as people in general (ignoring gender) take more of the drug, their blood pressure levels decrease (pictured below). We would have missed the fact that the drug works very differently on men than it does on women. The effect that gender has on the relationship between the amount of the drug taken and blood pressure levels is an example of an interaction (specifically a "disordinal" interaction).



In this study, it is possible that there is a disordinal interaction at work which you should consider in making your predictions of job performance. **Specifically, it is possible that what the cognitive ability test score tells you about an applicant's job performance (if the applicant were hired) depends upon how intellectually interesting or boring the job is.**

Appendix F

Instructions to Judgment Task (Spring 2010)

Prize:
The participant in this study who, on average, predicts job performance of the job applicants better than every other participant does will win $75 cash. The participant who performs second-best will win $50 cash, and the participant who performs third-best will win $25 cash. Participants who win will be contacted when this study is completed, so please be sure that we have contact information that will still be valid in August 2010.

Introduction:
Imagine that you are a hiring manager responsible for predicting the job performance of 125 job applicants. All applicants are applying for the same job. There are 2 key facts to consider when judging how well you expect each applicant to perform.

1. Cognitive Ability Test Score:
You will be told how each job applicant scored on a recent cognitive ability (intelligence) test. All applicants took the same exact test. Research shows a moderate to strong relationship between cognitive ability and job performance (e.g., Schmidt & Hunter, 1998). An applicant's cognitive ability test results are scored on a percentile between 0 to 100, but scores lower than the 1st percentile and greater than the 99th percentile are uncommon (see table below). A higher percentile score indicates a higher level of cognitive ability. A lower percentile score indicates a lower level of cognitive ability. The following table estimates some of the probabilities that a job applicant chosen randomly will have received certain percentile scores on the cognitive ability test:

	Percentile	Explanation of Percentile
Higher Score	99.99th	1 out of 10,000 applicants scores higher
↑	99.9th	1 out of 1,000 applicants scores higher
↑	99th	1 out of 100 applicants scores higher
↑	95th	1 out of 20 applicants scores higher
↑	80th	1 out of 5 applicants scores higher
↑	65th	7 out of 20 applicants score higher
---	50th	almost half of applicants score higher, and almost half score lower
↓	35th	7 out of 20 applicants score lower
↓	20th	1 out of 5 applicants scores lower
↓	5th	1 out of 20 applicants scores lower
↓	1st	1 out of 100 applicants scores lower
↓	.1th	1 out of 1,000 applicants scores lower
Lower Score	.01th	1 out of 10,000 applicants scores lower

2. Whether the Applicant Is Expected to Find the Job Interesting or Boring:
Job applicants are applying for the same exact job. For each job applicant, you will be told whether the applicant is expected to find the job interesting or boring based on the applicant's score on a job interest survey.

An applicant might apply for, accept, and continue working in a job that the applicant is expected to find boring because of many possible reasons (e.g., the applicant does not realize that he or she would find the job boring unless the applicant were to work in the job for several months, and by that point the applicant might feel committed to continue working in it; or the job market is bad, the job pays very well, and the applicant needs the income; etc.).

Your Task:
Your task is to predict how each job applicant would perform in the job if hired. People can be complex, so know that perfect prediction is extremely unlikely. Just do your best. You are to rate each applicant on a percentile from 0 to 100. In the past, everybody who applied for the job was hired and performed the job, but not every current and future applicant will be hired. It has been uncommon for job applicants to receive job performance ratings lower than the 1st percentile or greater than the 99th percentile (see table below). A higher percentile rating indicates that a job applicant would perform better in the job. A lower percentile rating indicates that a job applicant would perform worse in the job. The following table estimates some of the probabilities that a job applicant chosen randomly will have received certain job performance percentile ratings:

	Percentile	Explanation of Percentile
Higher Rating	99.99th	1 out of 10,000 applicants has received a higher rating
↑	99.9th	1 out of 1,000 applicants has received a higher rating
↑	99th	1 out of 100 applicants has received a higher rating
↑	95th	1 out of 20 applicants has received a higher rating
↑	80th	1 out of 5 applicants has received a higher rating
↑	65th	7 out of 20 applicants have received a higher rating
---	50th	almost half of applicants have received a higher rating, and almost half have received a lower rating
↓	35th	7 out of 20 applicants have received a lower rating
↓	20th	1 out of 5 applicants has received a lower rating
↓	5th	1 out of 20 applicants has received a lower rating
↓	1st	1 out of 100 applicants has received a lower rating
↓	.1th	1 out of 1,000 applicants has received a lower rating
Lower Rating	.01th	1 out of 10,000 applicants has received a lower rating

For each job applicant, please indicate your rating clearly and neatly in the space provided.

You have received files for the 125 job applicants. There is 1 file for each job applicant. A file contains a combination of cognitive ability test score and results from a job interest survey which indicate whether the applicant is expected to find the job interesting or boring.

Feedback:
After every set of 25 files (except for the very last set), you will receive the feedback information about how job performance relates to cognitive ability test score. Feedback information will appear as illustrations (i.e., pictures) followed by an explanation to help you interpret the illustrations.

1. Functional Form: This information will tell you something about how job performance relates to cognitive ability test score.
2. Predictability: This information tells you how well a pre-existing mathematical formula would have predicted job performance.
3. Relative Importance: This information tells you which of the following is more important to predicting job performance: (a) cognitive ability test score or (b) whether the applicant is expected to find the job interesting or boring. Assuming that one of them will be more important than the other, this information also will tell you by how much one of them is more important than the other.
4. Redundancy: This information tells you the extent to which (a) cognitive ability test score and (b) whether the applicant is expected to find the job interesting or boring tell you the same exact thing about how well the job applicant will perform in the job. In making a prediction, a hiring manager sometimes will use pieces of information that are redundant to some extent. Usually, the more that those pieces of information independently tell you something about an outcome (e.g., job performance), the more useful those pieces of information can be.

If you have any questions, please ask the person leading the study.

Appendix G

Sample of Prediction Profiles Form (Spring 2010)

Below are the cognitive ability test scores for each of 25 job applicants.
For each applicant it is noted below whether the applicant is expected to find the job interesting or boring.
In the space provided, please indicate your prediction of each applicant's job performance if the applicant were hired.
Below at right are the rating scales that you should use for job performance and cognitive ability.
Please do not leave this page until you have made predictions for all 25 job applicants.

Job applicant's identification number:	**Job applicant's cognitive ability percentile score (between 0 and 100):**	**Is the applicant expected to find the job interesting or boring?:**	**Your prediction of the applicant's job performance percentile if the applicant were hired? (between 0 and 100):**
51	92nd	Boring	
52	88th	Interesting	
53	42nd	Boring	
54	21st	Boring	
55	69th	Interesting	
56	1st	Interesting	
57	38th	Boring	
58	50th	Interesting	
59	54th	Interesting	
60	12th	Interesting	
61	34th	Interesting	
62	76th	Boring	
63	69th	Interesting	
64	90th	Boring	
65	58th	Boring	
66	34th	Boring	
67	50th	Boring	
68	34th	Boring	
69	54th	Interesting	
70	58th	Boring	
71	79th	Boring	
72	4th	Boring	
73	96th	Boring	
74	58th	Interesting	
75	73rd	Boring	

Job Performance Percentile Rating Scale:

	Percentile	Explanation of Percentile
Higher Rating	99.99th	1 out of 10,000 applicants has received a higher rating
↑	99.9th	1 out of 1,000 applicants has received a higher rating
↑	99th	1 out of 100 applicants has received a higher rating
↑	95th	1 out of 20 applicants has received a higher rating
↑	80th	1 out of 5 applicants has received a higher rating
↑	65th	7 out of 20 applicants have received a higher rating
---	50th	almost half of applicants have received a higher rating, and almost half have received a lower rating
↓	35th	7 out of 20 applicants have received a lower rating
↓	20th	1 out of 5 applicants has received a lower rating
↓	5th	1 out of 20 applicants has received a lower rating
↓	1st	1 out of 100 applicants has received a lower rating
↓	.1th	1 out of 1,000 applicants has received a lower rating
Lower Rating	.01th	1 out of 10,000 applicants has received a lower rating

Cognitive Ability Test Rating Scale:

	Percentile	Explanation of Percentile
Higher Score	99.99th	1 out of 10,000 applicants scores higher
↑	99.9th	1 out of 1,000 applicants scores higher
↑	99th	1 out of 100 applicants scores higher
↑	95th	1 out of 20 applicants scores higher
↑	80th	1 out of 5 applicants scores higher
↑	65th	7 out of 20 applicants score higher
---	50th	almost half of applicants score higher, almost half lower
↓	35th	7 out of 20 applicants score lower
↓	20th	1 out of 5 applicants scores lower
↓	5th	1 out of 20 applicants scores lower
↓	1st	1 out of 100 applicants scores lower
↓	.1th	1 out of 1,000 applicants scores lower
Lower Score	.01th	1 out of 10,000 applicants scores lower

A-C1-3 of 5

Appendix H

Major Differences Among the Feedback Treatment Conditions (Spring 2010) (See Appendix I for the Graphical Feedback Itself)

	Condition 1	**Condition 2**	**Condition 3**
1. Purpose	To determine the extent to which subjects can identify and effectively utilize the disordinal interaction without being explicitly told that it might exist or does exist. (However, as feedback, a scatterplot is provided to describe the relationship between cognitive ability and job performance for interesting jobs and then separately and adjacently for boring jobs.)	To determine the extent to which subjects can identify and effectively utilize the disordinal interaction if explicitly told that it might exist. (As feedback, the scatterplots from Condition 1 appear for Condition 2 along but with a detailed example of a disordinal interaction for an unrelated context.)	To determine the extent to which subjects can identify and effectively utilize the disordinal interaction if explicitly told that it does exist. (The disordinal interaction is explicitly illustrated and explained. The feedback is intended as clear evidence that the interaction needs to be used.)
2. Feedback: Scatterplots and Line Graphs	Scatterplots (Y = actual job performance; X = cognitive ability test score). One plot for interesting jobs, another for boring jobs. No best-fitting lines appear. Data describe immediately prior block of predictions made. No line graphs	Same scatterplots as in Condition 1. Detailed example of a disordinal interaction for an unrelated context. No line graphs.	Same scatterplots as in Condition 1. Same detailed example of disordinal interaction as in Condition 2. Line graph explicitly showing the disordinal interaction (Y = actual job performance; X = cognitive ability test score). There are 2 separate lines: one for when the applicant is expected to find the job interesting and one for when the applicant is expected to find the job boring. Depicts all 10,000 cases from the original data simulation (the population).

	Condition 1	Condition 2	Condition 3
3. Feedback: Pie Charts	Pie charts: Provide task information feedback without the disordinal interaction. Feedback includes: 1. How predictable job performance is when using a linear mechanical equation <u>(with nothing about predictability for the interaction mechanical equation)</u> 2. C_{xj} relative weights (dominance analysis) for (1) cognitive ability test score and (2) whether the applicant is expected to find the job interesting or boring <u>(with nothing about the interaction term)</u> 3. Redundancy (inter-correlation) for (1) cognitive ability test score and (2) whether the applicant is expected to find the job interesting or boring.	Same as Condition 1.	Pie charts: Designed to emphasize importance of the disordinal interaction. Feedback includes: 1. How predictable job performance is when using a linear mechanical equation <u>versus how predictable job performance is when using an interaction mechanical equation</u> 2. C_{xj} relative weights (dominance analysis) for (1) cognitive ability test score, (2) whether the applicant is expected to find the job interesting or boring, <u>and (3) the interaction term</u>. Note: Cue redundancy information is unnecessary if the goal is to explicitly promote use of the interaction, so that information is excluded.

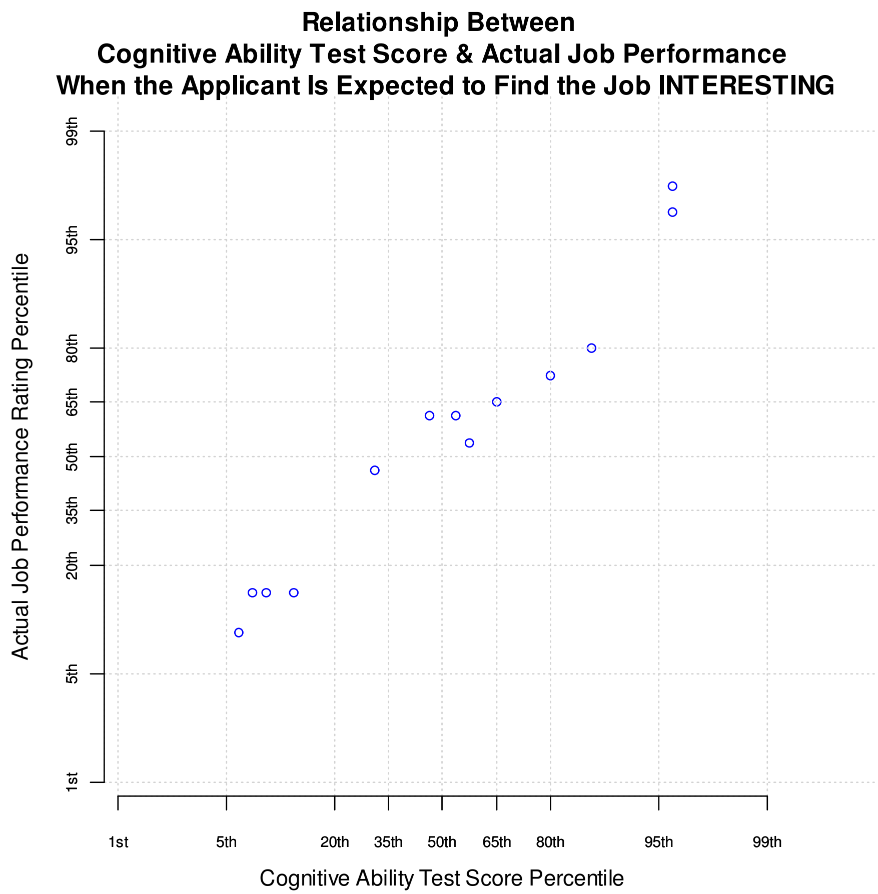
	Condition 1	Condition 2	Condition 3
4. Subjects' narrative self-reports of confidence level and strategies used	No explicit mention at all of any disordinal interaction; identification and/or use of it might be captured in the answer to one of the following 3 items/questions: 1. Please describe your strategies that you used to make the previous 25 predictions of job performance: 2. What do you think the relationship is between cognitive ability test score and job performance? 3. In making the previous 25 predictions, did you try to use information <u>other than or in addition to</u> cognitive ability test score or whether the applicant is expected to find the job interesting or boring?	Designed to ask explicitly about whether the disordinal interaction was used	Same as in Condition 2.

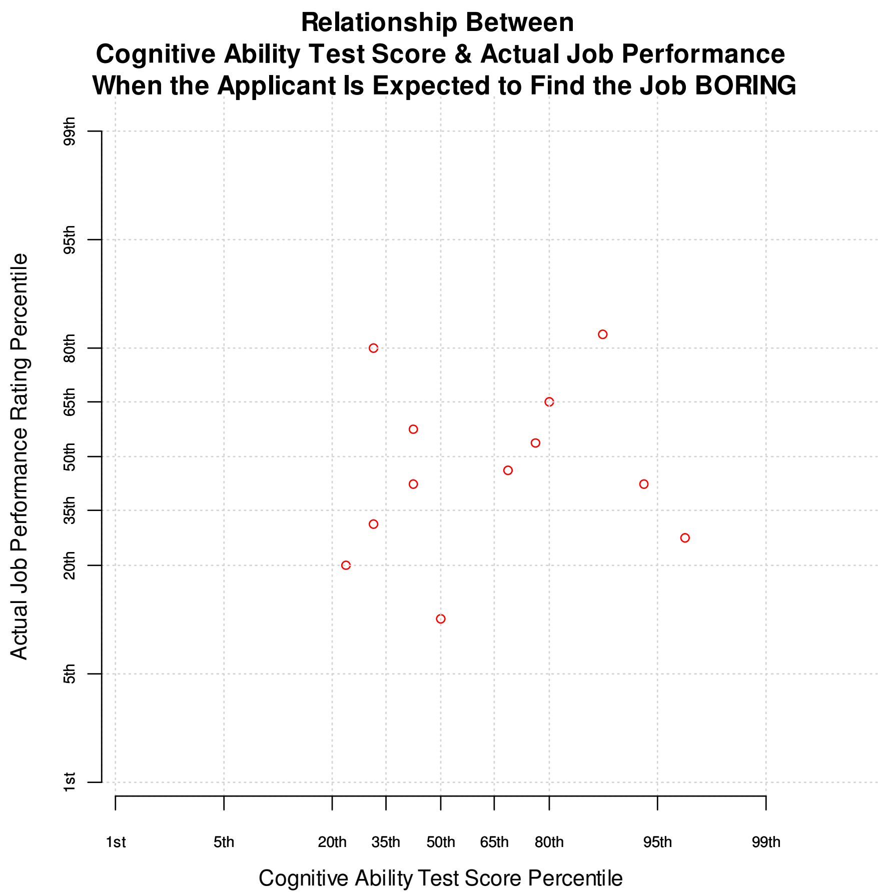
Appendix I

Task Information Feedback (Spring 2010)
(See Appendix H for a Narrative Description of Major Differences Among the Feedback Treatment Conditions)

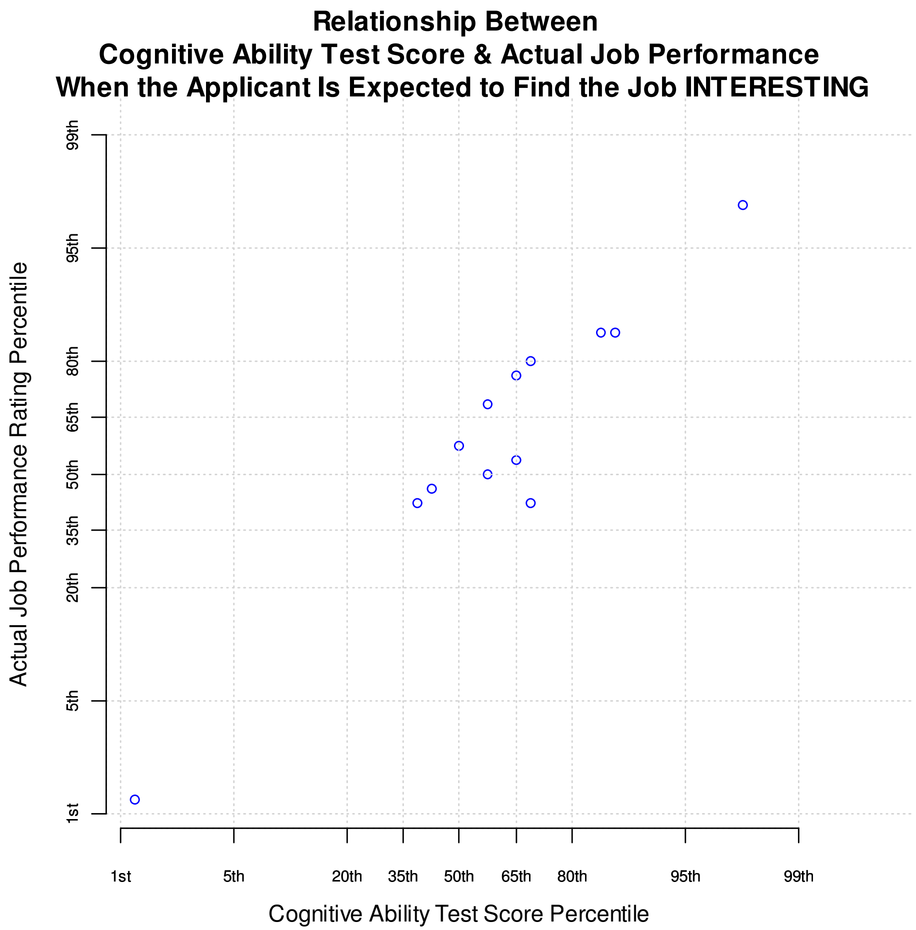
Scatterplots for Each Block (All 3 Feedback Conditions)

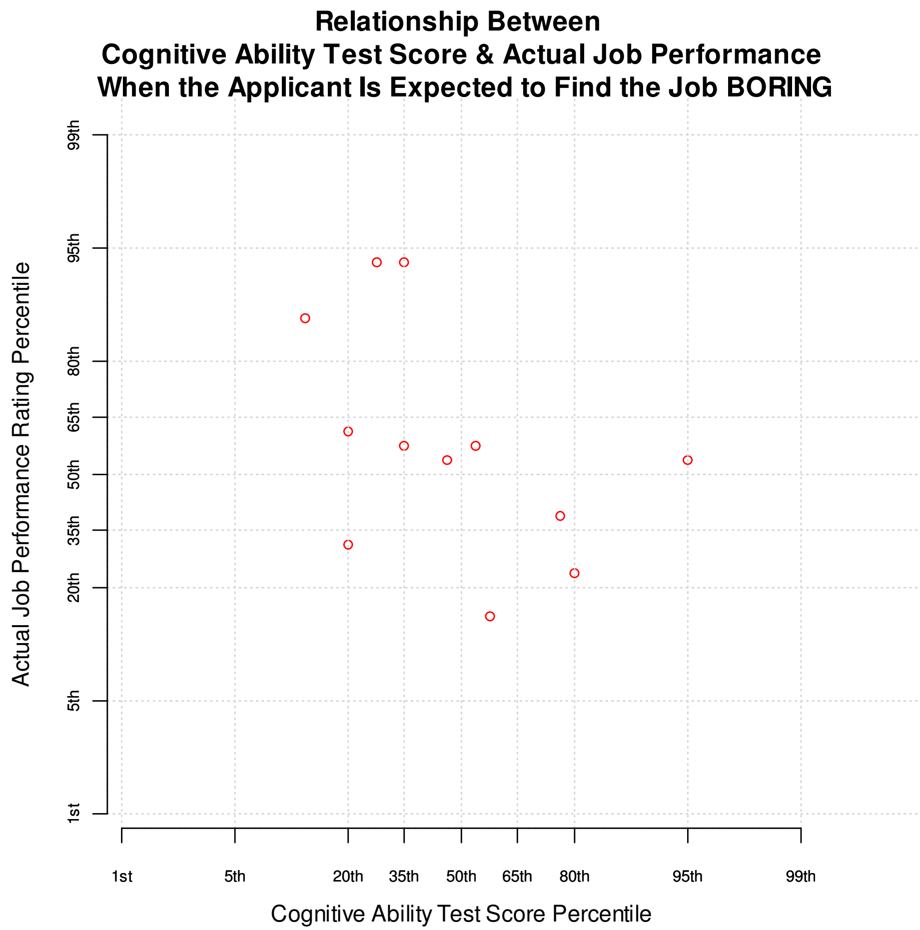
Block 1:





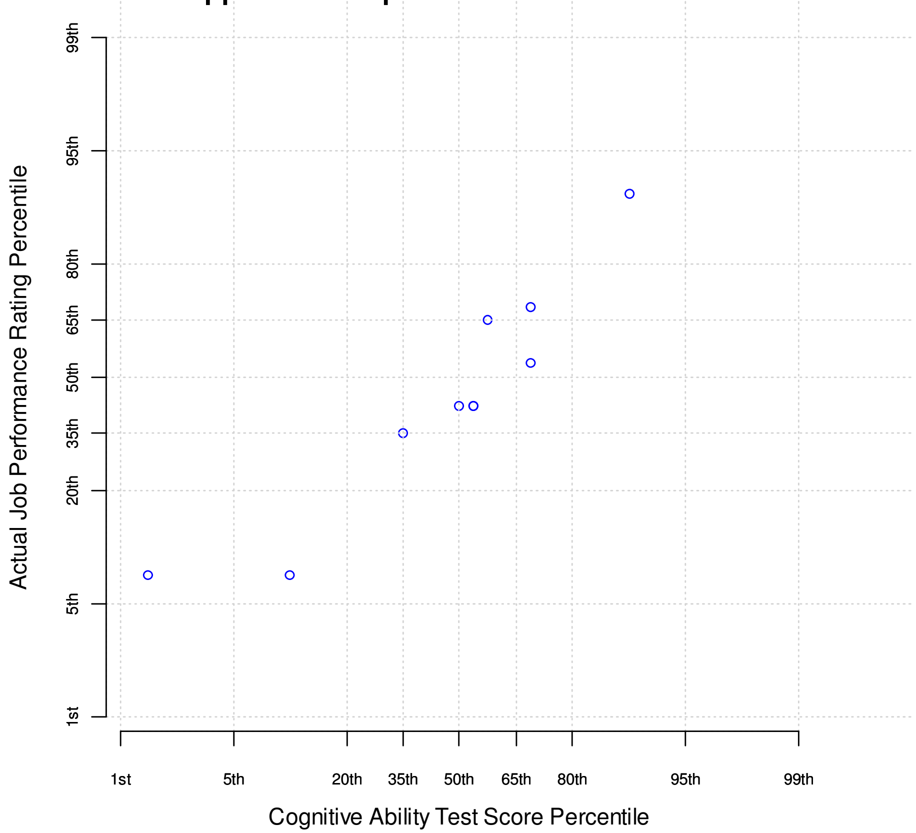
Block 2:


Relationship Between
Cognitive Ability Test Score & Actual Job Performance
When the Applicant Is Expected to Find the Job INTERESTING
Actual Job Performance Rating Percentile
99th
95th
80th
65th
50th
35th
20th
5th
1st
1st
5th
20th
35th
50th
65th
80th
95th
99th
Cognitive Ability Test Score Percentile

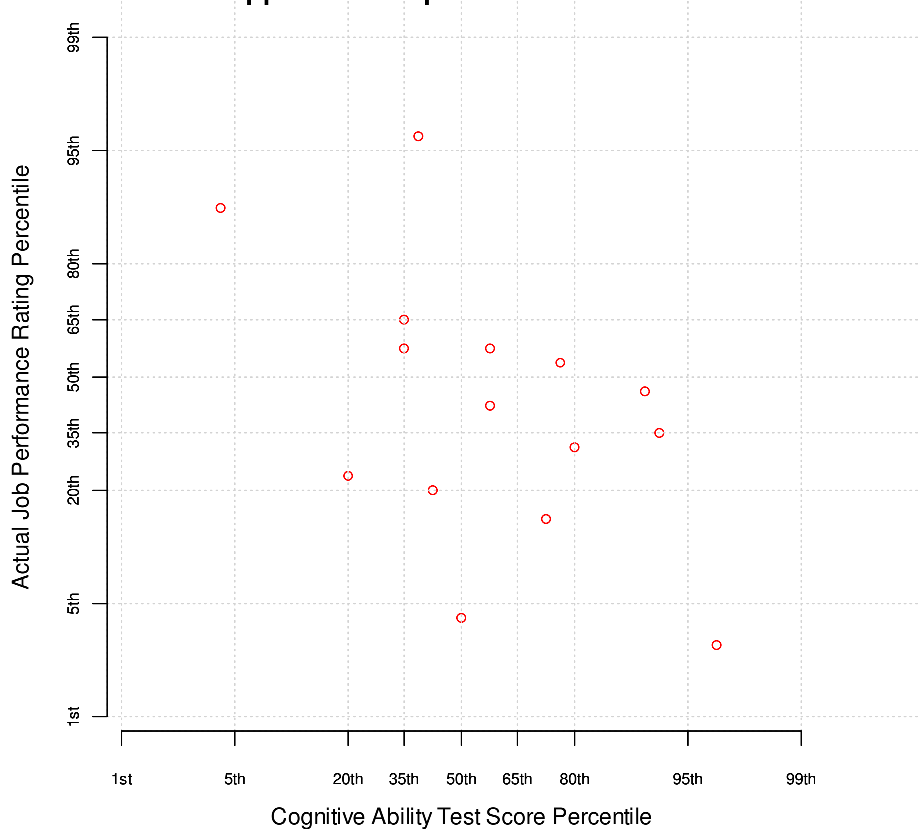

Relationship Between
Cognitive Ability Test Score & Actual Job Performance
When the Applicant Is Expected to Find the Job BORING
Actual Job Performance Rating Percentile
99th
95th
80th
65th
50th
35th
20th
5th
1st
1st
5th
20th
35th
50th
65th
80th
95th
99th
Cognitive Ability Test Score Percentile

Block 3:

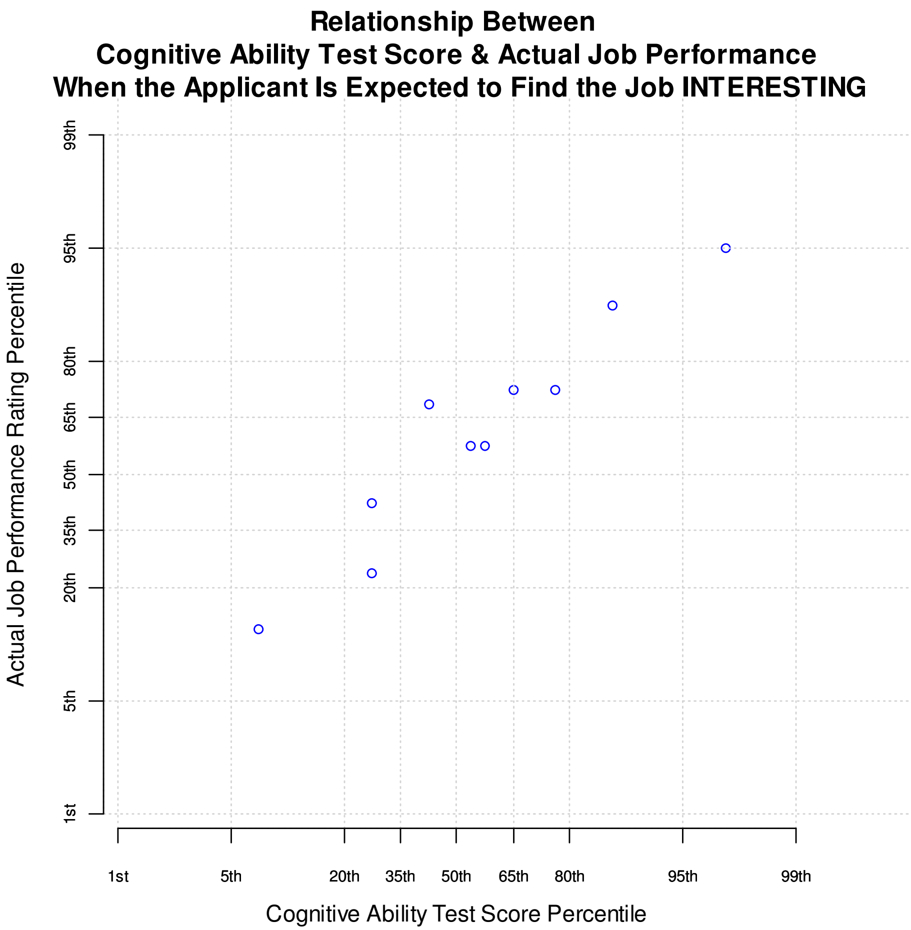
Relationship Between Cognitive Ability Test Score & Actual Job Performance When the Applicant Is Expected to Find the Job INTERESTING

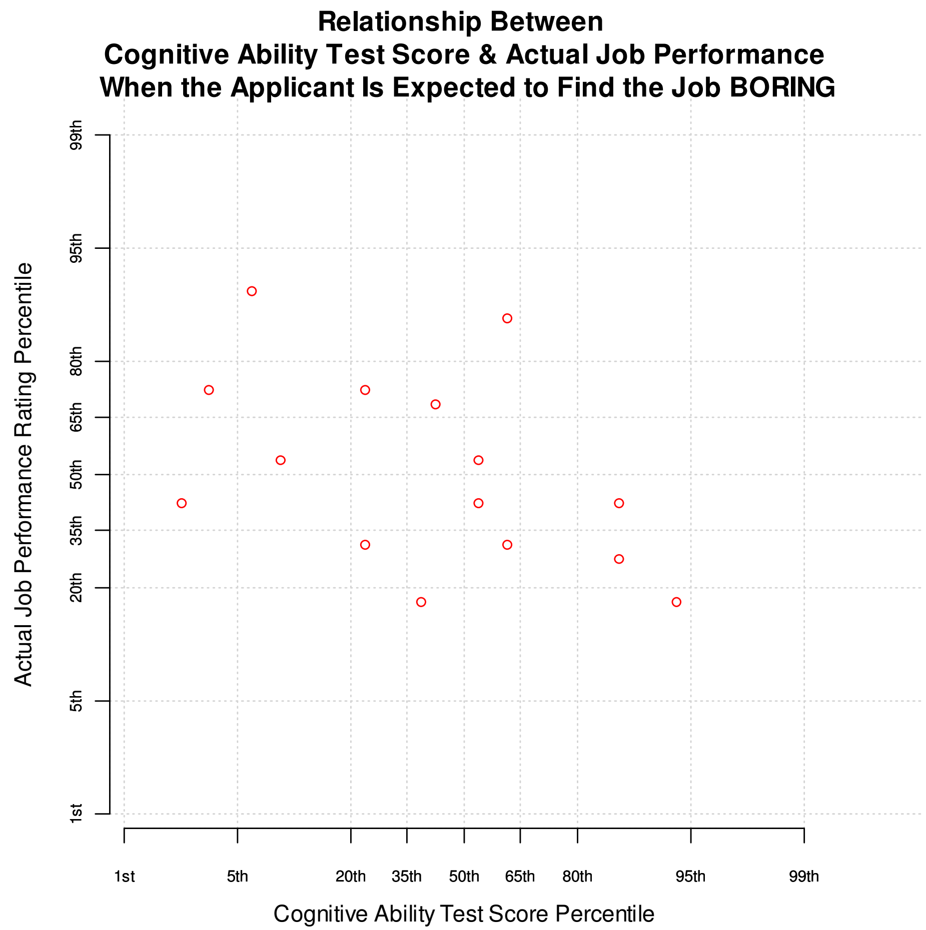


Relationship Between Cognitive Ability Test Score & Actual Job Performance When the Applicant Is Expected to Find the Job BORING

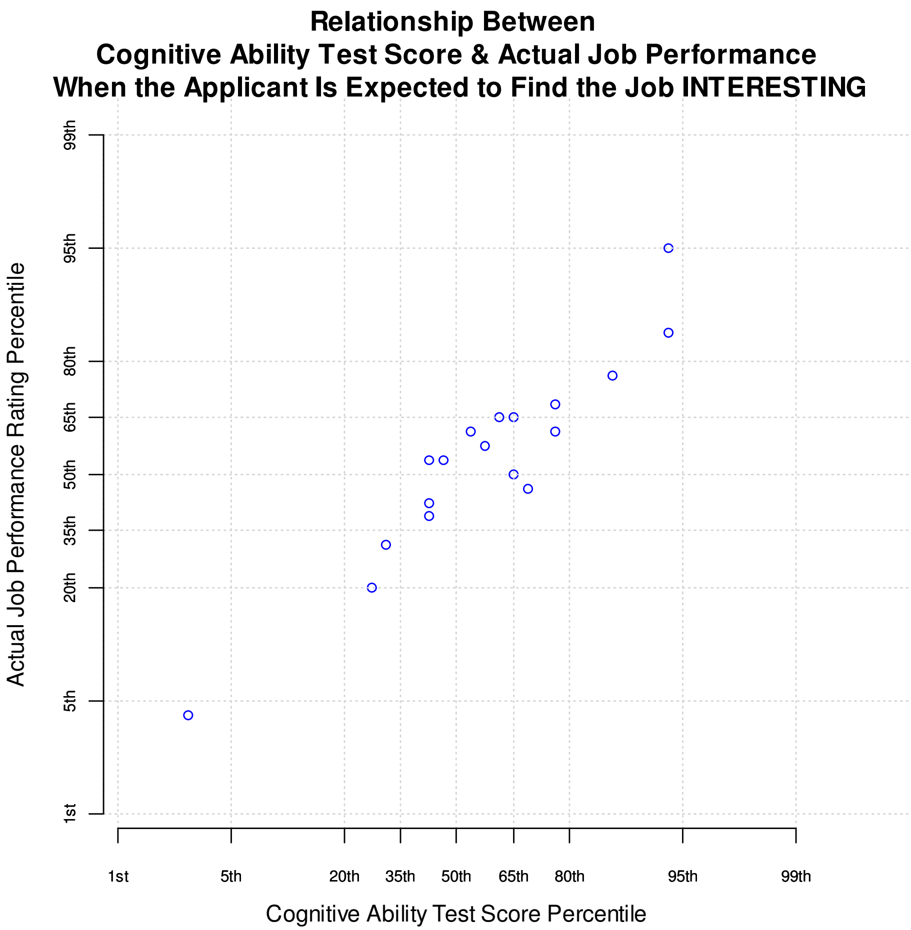


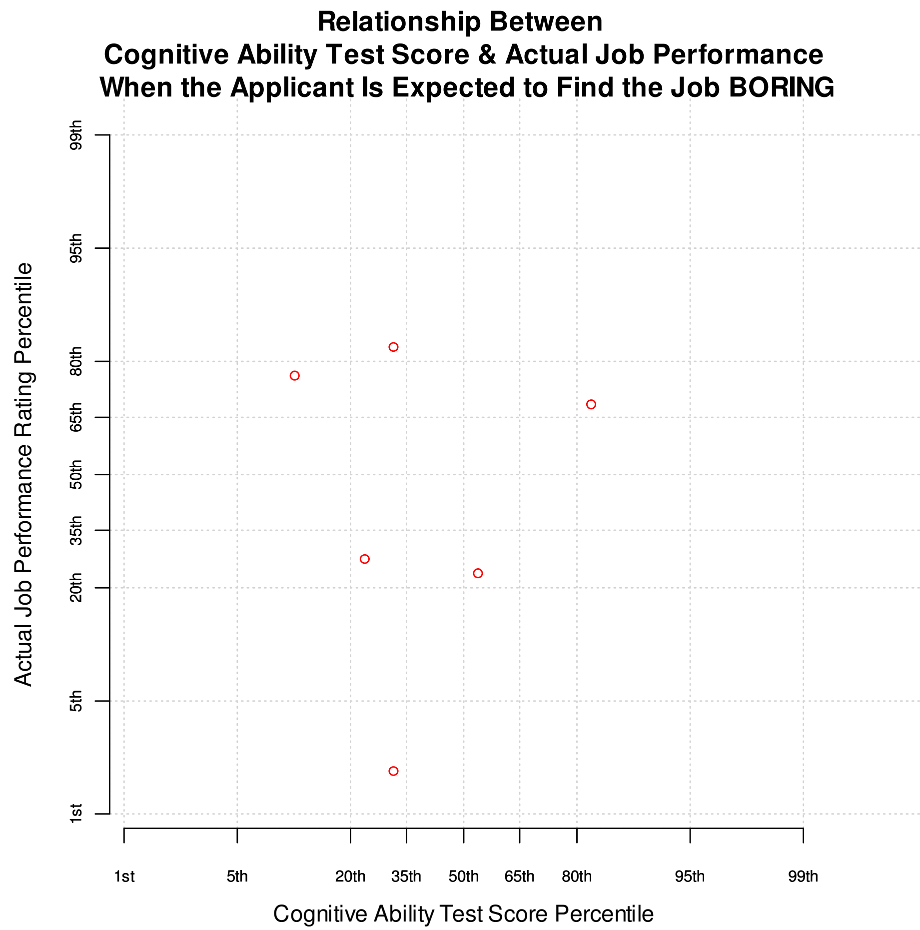
Block 4:


Relationship Between
Cognitive Ability Test Score & Actual Job Performance
When the Applicant Is Expected to Find the Job INTERESTING
Actual Job Performance Rating Percentile
99th
95th
80th
65th
50th
35th
20th
5th
1st
1st
5th
20th
35th
50th
65th
80th
95th
99th
Cognitive Ability Test Score Percentile


Relationship Between
Cognitive Ability Test Score & Actual Job Performance
When the Applicant Is Expected to Find the Job BORING
Actual Job Performance Rating Percentile
99th
95th
80th
65th
50th
35th
20th
5th
1st
1st
5th
20th
35th
50th
65th
80th
95th
99th
Cognitive Ability Test Score Percentile

Block 5:


Relationship Between
Cognitive Ability Test Score & Actual Job Performance
When the Applicant Is Expected to Find the Job INTERESTING
Actual Job Performance Rating Percentile
99th
95th
80th
65th
50th
35th
20th
5th
1st
1st
5th
20th
35th
50th
65th
80th
95th
99th
Cognitive Ability Test Score Percentile

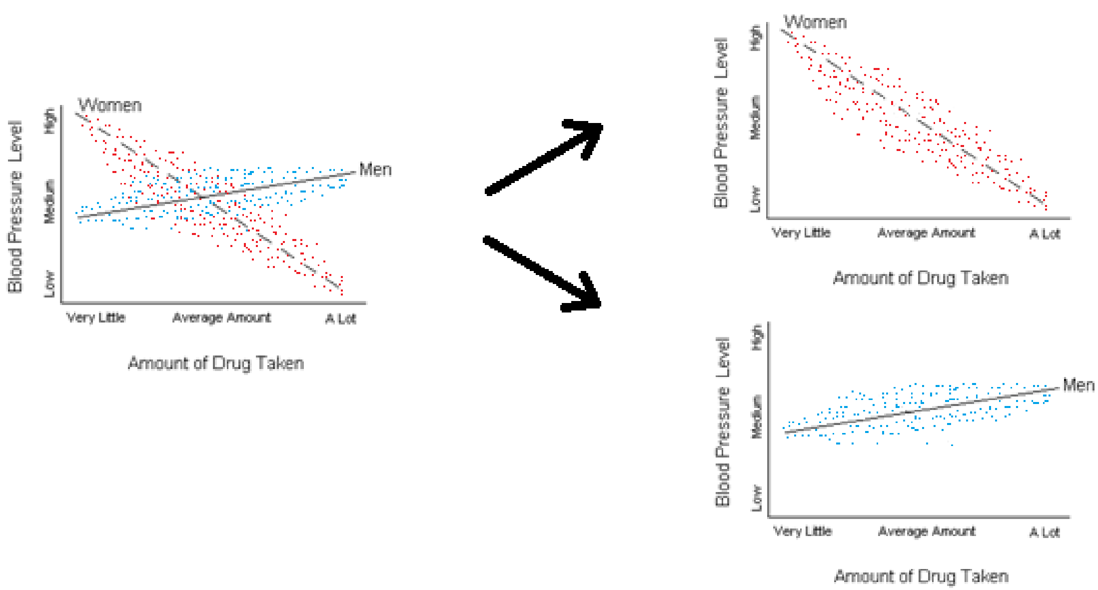

Relationship Between
Cognitive Ability Test Score & Actual Job Performance
When the Applicant Is Expected to Find the Job BORING
Actual Job Performance Rating Percentile
99th
95th
80th
65th
50th
35th
20th
5th
1st
1st
5th
20th
35th
50th
65th
80th
95th
99th
Cognitive Ability Test Score Percentile

<u>How to interpret the illustrations on the previous page</u>:

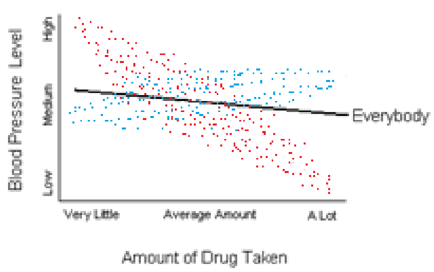
Pictured on the previous page for the immediately previous 25 job applicants is the relationship between cognitive ability test score and actual job performance (a) for those applicants expected to find the job interesting and (b) for those applicants expected to find the job boring. Each of the 25 applicants is represented by a single circular point. You might see one or more trends in the scatters of the points.

As in other situations where predictions are made, it is possible that what one piece of information tells you about an outcome depends upon the value of another piece of information. This dependency is called an "interaction". One example of an interaction (specifically a "disordinal" interaction) is depicted below, at left. It shows that as women take more of a drug to keep blood pressure levels low, their blood pressure levels drop. However, as men take more of the same drug, their blood pressure levels increase. (Lines indicating a different trend for each gender are included in addition to the points.)

We can take the picture at left and break it into 2 separate pictures, one for just women and one for just men:



What if we had ignored gender and just looked at everybody together? We would have seen that as people in general (ignoring gender) take more of the drug, their blood pressure levels decrease (pictured below with a single overall trend line in addition to the points). We would have missed the fact that the drug works very differently on men than it does on women. The effect that gender has on the relationship between the amount of the drug taken and blood pressure levels is an example of an interaction (specifically a "disordinal" interaction).



{When looking at relationships, it is possible that there is a disordinal interaction at work which you should consider when making predictions. Whether or not one exists in this study is for you to decide.}← LANGUAGE USED FOR CONDITION 2 ONLY

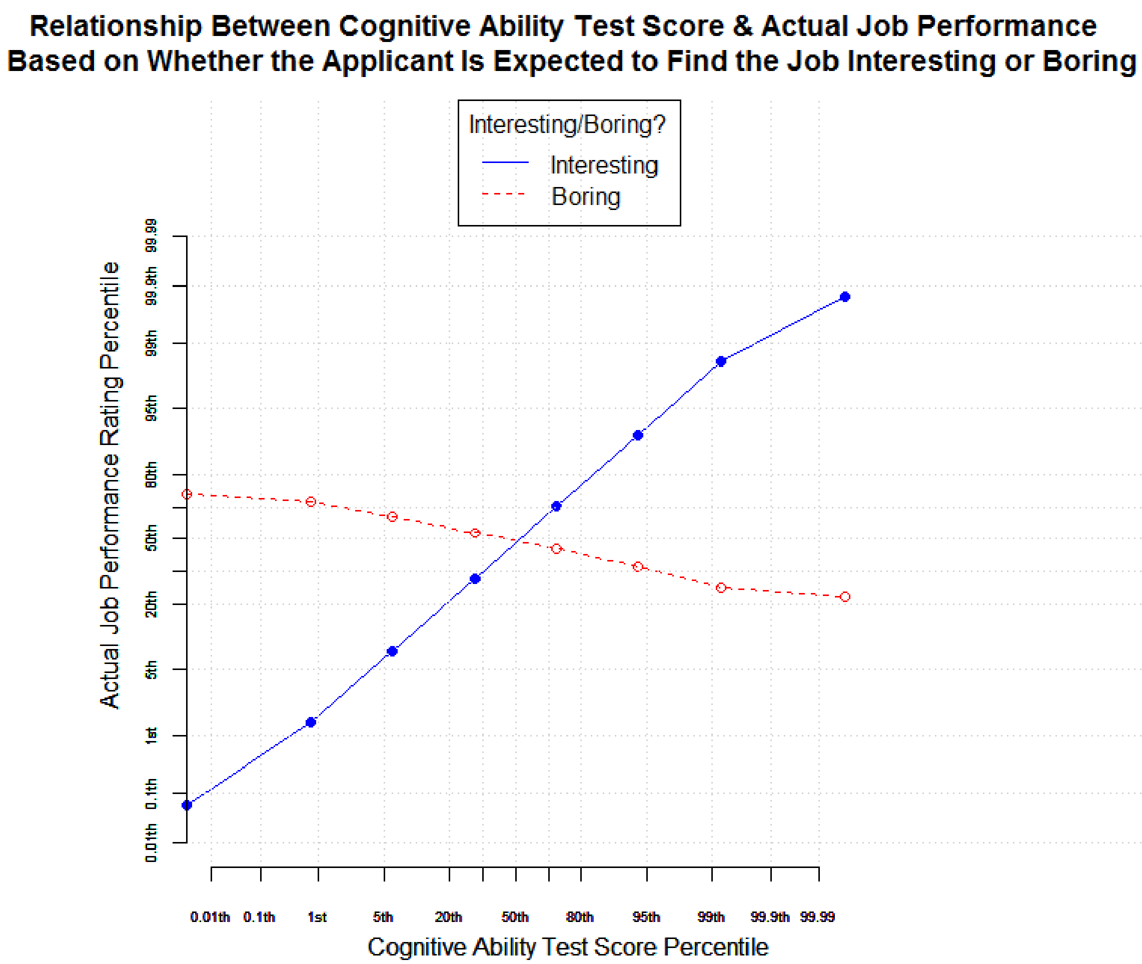
<u>{In this study, there is a disordinal interaction at work which you should consider when making predictions of job performance.</u>}← LANGUAGE USED FOR CONDITION 3 ONLY

In other words, the relationship between cognitive ability test score and actual job performance depends upon whether the applicant is expected to find the job interesting or boring.

If the applicant is expected to find the job interesting, then the relationship between cognitive ability test score and actual job performance is very strong and positive, and you should predict better (higher) job performance for a higher cognitive ability test score. If an applicant finds a job to be interesting, then greater intelligence will almost always lead to better job performance.

If the applicant is expected to find the job boring, then the relationship between cognitive ability test score and actual job performance is negative (although very weak), and you should predict lower job performance for a higher cognitive ability test score. If an applicant finds a job to be boring, then greater intelligence usually will lead to greater boredom and lower job performance.

Below is an another illustration of the relationship between cognitive ability and actual job performance. However, this is an illustration of the disordinal interaction when we look at the thousands of people who applied for the job in the past, and not simply the immediately previous 25 applicants currently applying for the job. (In the past, everybody who applied for the job was hired and performed the job, but not every current and future applicant will be hired.) All of the lines and circular points in the illustration below are based on averaging across applicants. Generally, the average (which is indicated by the lines and circular points in the picture below) will be your best guess of job performance.



Additional Feedback – All Feedback Condiitons (Subject to Color Coding)

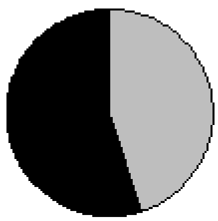
The following feedback information is related to the immediately previous 25 job applicants whose job performance you most recently predicted:

Block 1:

Red text or frame = information that appears for Feedback Condition 3 only

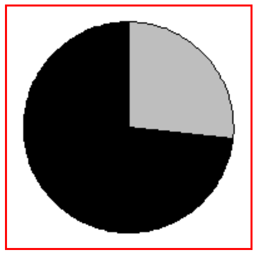
Blue text or frame = information that appears for Feedback Conditions 1 and 2 only

How predictable was job performance?



if a prediction equation did not consider the disordinal interaction

if a prediction equation considered only (1) cognitive ability test score and (2) whether the applicant is expected to find the job interesting or boring



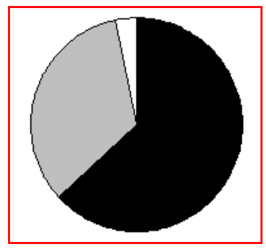
if a prediction equation did consider the disordinal interaction

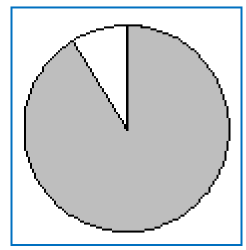
Total pie = perfect prediction (in an ideal world)

Black section = the actual predictability when using an equation created in advance

The bigger the black section, the more predictable that job performance is.

How important to predicting actual job performance was (1) the cognitive ability test score versus (2) whether the applicant is expected to find the job interesting or boring versus (3) the disordinal interaction.





Total pie = everything that should be considered when predicting job performance

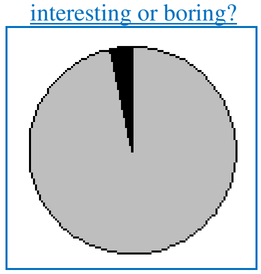
Gray section = cognitive ability test score

White section = whether the applicant is expected to find the job interesting or boring

Black section = the disordinal interaction

The bigger the section, the more important that it is to predicting job performance.

How redundant was (1) the cognitive ability test score to (2) whether the applicant is expected to find the job interesting or boring?



Total pie = total redundancy (i.e., if these two pieces of information told you the same exact thing about job performance)

Black section = the actual extent of redundancy

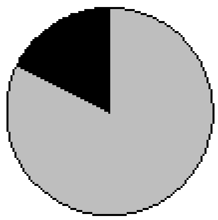
The bigger the black section, the more that these two pieces of information tell you the same exact thing about job performance.

Block 2:

Red text or frame = information that appears for Feedback Condition 3 only

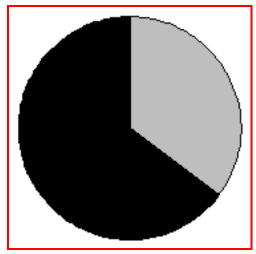
Blue text or frame = information that appears for Feedback Conditions 1 and 2 only

How predictable was job performance?



if a prediction equation did not consider the disordinal interaction

if a prediction equation considered only (1) cognitive ability test score and (2) whether the applicant is expected to find the job interesting or boring



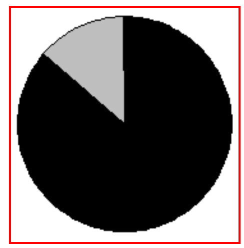
if a prediction equation did consider the disordinal interaction

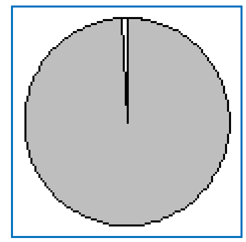
Total pie = perfect prediction (in an ideal world)

Black section = the actual predictability when using an equation created in advance

The bigger the black section, the more predictable that job performance is.

How important to predicting actual job performance was (1) the cognitive ability test score versus (2) whether the applicant is expected to find the job interesting or boring versus (3) the disordinal interaction.





Total pie = everything that should be considered when predicting job performance

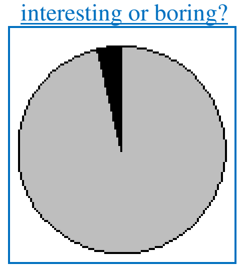
Gray section = cognitive ability test score

White section = whether the applicant is expected to find the job interesting or boring

Black section = the disordinal interaction

The bigger the section, the more important that it is to predicting job performance.

How redundant was (1) the cognitive ability test score to (2) whether the applicant is expected to find the job interesting or boring?



Total pie = total redundancy (i.e., if these two pieces of information told you the same exact thing about job performance)

Black section = the actual extent of redundancy

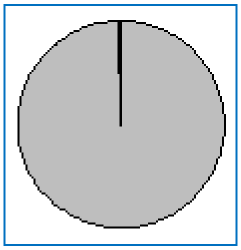
The bigger the black section, the more that these two pieces of information tell you the same exact thing about job performance.

Block 3:

Red text or frame = information that appears for Feedback Condition 3 only

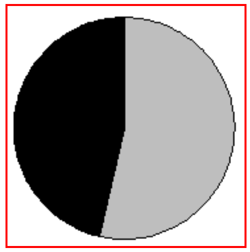
Blue text or frame = information that appears for Feedback Conditions 1 and 2 only

How predictable was job performance?



if a prediction equation did not consider the disordinal interaction

if a prediction equation considered only (1) cognitive ability test score and (2) whether the applicant is expected to find the job interesting or boring



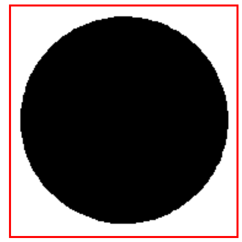
if a prediction equation did consider the disordinal interaction

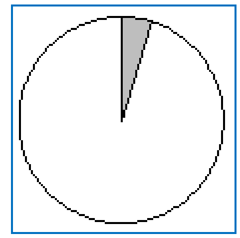
Total pie = perfect prediction (in an ideal world)

Black section = the actual predictability when using an equation created in advance

The bigger the black section, the more predictable that job performance is.

How important to predicting actual job performance was (1) the cognitive ability test score versus (2) whether the applicant is expected to find the job interesting or boring versus (3) the disordinal interaction.





Total pie = everything that should be considered when predicting job performance

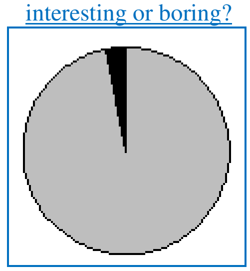
Gray section = cognitive ability test score

White section = whether the applicant is expected to find the job interesting or boring

Black section = the disordinal interaction

The bigger the section, the more important that it is to predicting job performance.

How redundant was (1) the cognitive ability test score to (2) whether the applicant is expected to find the job interesting or boring?



Total pie = total redundancy (i.e., if these two pieces of information told you the same exact thing about job performance)

Black section = the actual extent of redundancy

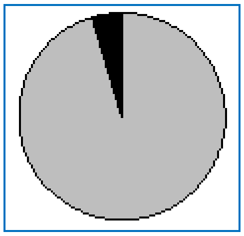
The bigger the black section, the more that these two pieces of information tell you the same exact thing about job performance.

Block 4:

Red text or frame = information that appears for Feedback Condition 3 only

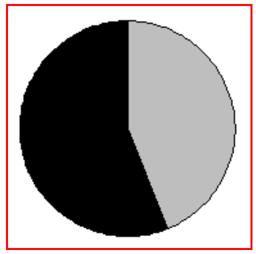
Blue text or frame = information that appears for Feedback Conditions 1 and 2 only

How predictable was job performance?



if a prediction equation did not consider the disordinal interaction

if a prediction equation considered only (1) cognitive ability test score and (2) whether the applicant is expected to find the job interesting or boring



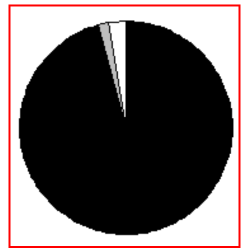
if a prediction equation did consider the disordinal interaction

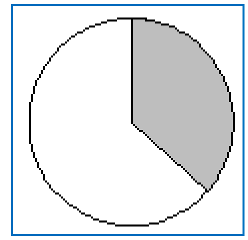
Total pie = perfect prediction (in an ideal world)

Black section = the actual predictability when using an equation created in advance

The bigger the black section, the more predictable that job performance is.

How important to predicting actual job performance was (1) the cognitive ability test score versus (2) whether the applicant is expected to find the job interesting or boring versus (3) the disordinal interaction.





Total pie = everything that should be considered when predicting job performance

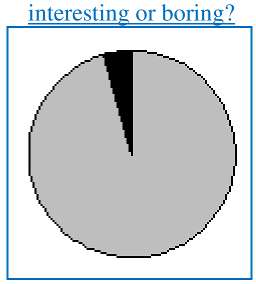
Gray section = cognitive ability test score

White section = whether the applicant is expected to find the job interesting or boring

Black section = the disordinal interaction

The bigger the section, the more important that it is to predicting job performance.

How redundant was (1) the cognitive ability test score to (2) whether the applicant is expected to find the job interesting or boring?



Total pie = total redundancy (i.e., if these two pieces of information told you the same exact thing about job performance)

Black section = the actual extent of redundancy

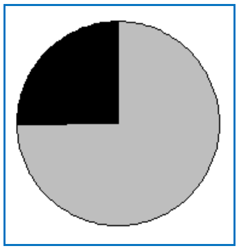
The bigger the black section, the more that these two pieces of information tell you the same exact thing about job performance.

Block 5:

Red text or frame = information that appears for Feedback Condition 3 only

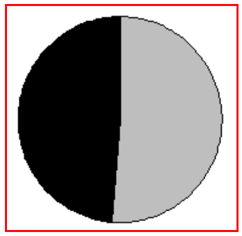
Blue text or frame = information that appears for Feedback Conditions 1 and 2 only

How predictable was job performance?



if a prediction equation did not consider the disordinal interaction

if a prediction equation considered only (1) cognitive ability test score and (2) whether the applicant is expected to find the job interesting or boring



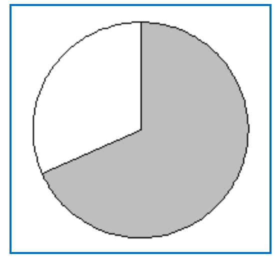
if a prediction equation did consider the disordinal interaction

Total pie = perfect prediction (in an ideal world)

Black section = the actual predictability when using an equation created in advance

The bigger the black section, the more predictable that job performance is.

How important to predicting actual job performance was (1) the cognitive ability test score versus (2) whether the applicant is expected to find the job interesting or boring versus (3) the disordinal interaction.



Total pie = everything that should be considered when predicting job performance

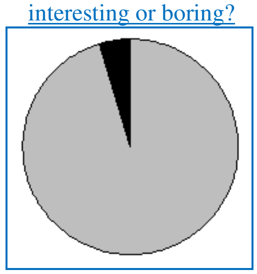
Gray section = cognitive ability test score

White section = whether the applicant is expected to find the job interesting or boring

Black section = the disordinal interaction

The bigger the section, the more important that it is to predicting job performance.

How redundant was (1) the cognitive ability test score to (2) whether the applicant is expected to find the job interesting or boring?



Total pie = total redundancy (i.e., if these two pieces of information told you the same exact thing about job performance)

Black section = the actual extent of redundancy

The bigger the black section, the more that these two pieces of information tell you the same exact thing about job performance.

Appendix J

Consent Form (Fall 2009)

Prediction of Performance

You are invited to be in a research study of the prediction of human performance. You were selected as a possible participant, because through the Psychology Department's research website, you indicated that you wished to participate. We ask that you read this form and ask any questions you may have before agreeing to be in the study.

This study is being conducted by: David M. Klieger, J.D., Department of Psychology, Ph.D. Candidate.

Background Information

The purpose of this study is to determine how well people predict others' performance if the people making the predictions are given certain information. In particular, this study is trying to ascertain whether these people can outperform a pre-existing equation that is being used to make the same predictions. The ultimate goal is to determine the extent to which people can learn to improve their ability to predict human behavior both absolutely and in comparison to pre-existing equations.

Procedures:

If you agree to be in this study, we would ask you to respond to a series of questions.

One set of questions (Academic Information & Experience Form) is designed to collect academic and experience information about you.

A second and third set of questions (Goldberg Quick Big-5 and IPC-7) are meant to collect information about your personality.

The third set of questions (Interest Profiler Form) tries to gather information about your interests.

The remaining sets of questions ask you to make predictions about the performance of hypothetical job applicants if they were hired.

The academic, experience, personality, and interest measures will help us to determine if people with certain characteristics perform better or worse than people with other characteristics in making the predictions. Because a major purpose of our study is to determine if academic background is related to performance on the prediction tasks, we will be asking your permission to allow the University of Minnesota to give us access to your college grade point average and number of completed academic hours as well as access to your high school percentile rank and standardized test (e.g., ACT, SAT) scores.

Risks and Benefits of being in the Study

ALL INFORMATION THAT YOU PROVIDE WILL BE KEPT CONFIDENTIAL, SECURED AND ANONYMOUS. As with any disclosure made to anyone, there always is a non-zero risk of unforeseeable breach of confidentiality, security and/or anonymity. However, this risk is minimal.

The benefits to participation are getting you to think broadly, precisely and carefully about how to make predictions about other people's behavior. Although some may take for granted the ability to make such predictions accurately, empirical research demonstrates that people often are not as accurate as they think they are. This overconfidence can lead to negative consequences, such as admitting to college a student who is not prepared and fails out of school, hiring an employee who disrupts a work unit, and advancing to CEO a reckless executive who bankrupts a company.

Compensation:

You will receive payment: 4 REP points shortly after we link your responses to the standardized test information from the university. If your predictions are, on average, more accurate than those of every other participant, then you will be paid $75 by the conclusion of this study. The participant who performs second-best will win $50, and the participant who performs third-best will win $25.

Confidentiality:

The records of this study will be kept private. In any sort of report we might publish, we will not include any information that will make it possible to identify a subject. Research records will be stored securely and only researchers will have access to the records.

Voluntary Nature of the Study:

Participation in this study is voluntary. Your decision whether or not to participate will not affect your current or future relations with the University of Minnesota. If you decide to participate, you are free to not answer any question or withdraw at any time with out affecting those relationships.

Contacts and Questions:

The researcher conducting this study is David M. Klieger, J.D. You may ask any questions you have now. If you have questions later, **you are encouraged** to contact David Klieger at The Department of Psychology, Elliott Hall, 75 East River Road, Minneapolis, MN 55455, (612) 625-3529, klie0019@umn.edu. You may contact his faculty advisor, Professor Nathan R. Kuncel at the same address, but at the following different telephone number and e-mail address: (612) 625-2818, kunce001@umn.edu.

If you have any questions or concerns regarding this study and would like to talk to someone other than the researcher(s), **you are encouraged** to contact the Research Subjects' Advocate Line, D528 Mayo, 420 Delaware St. Southeast, Minneapolis, Minnesota 55455; (612) 625-1650.

You will be given a copy of this information to keep for your records.

Statement of Consent:

I have read the above information. I have asked questions and have received answers. I consent to participate in the study and to the release by the University of Minnesota to the primary investigator, David M. Klieger, of my college grade point average(s), number of completed academic hours, high school rank, and standardized test scores (e.g., ACT, SAT) scores.

Signature:_____________________________ Date: __________________

Printed name: ______________________ University ID # __________________

Signature of parent or guardian:______________________ Date: __________________
(If minors are involved)

Printed name: ____________________________

Signature of Investigator:___________________ Date: __________________

Appendix K

Consent Form (Spring 2010)

Prediction of Performance

You are invited to be in a research study of the prediction of human performance. You were selected as a possible participant, because through the Psychology Department's research website, you indicated that you wished to participate. We ask that you read this form and ask any questions you may have before agreeing to be in the study.

This study is being conducted by: David M. Klieger, J.D., Department of Psychology, Ph.D. Candidate.

Background Information

The purpose of this study is to determine how well people predict others' performance if the people making the predictions are given certain information. In particular, this study is trying to ascertain whether people can outperform a pre-existing equation that is being used to make the same predictions. The ultimate goal is to determine the extent to which people can learn to improve their ability to predict human behavior both absolutely and in comparison to pre-existing equations.

Procedures

If you agree to be in this study, we would ask you to respond to a series of questions. The entire session will take 2 hours.

One set of questions (Academic Information & Experience Form) is designed to collect academic and experience information about you. (Example: "Describe your expectations for this study (e.g., if you expect to do well, average, or poorly; if you expect the predictions to be easy or hard to make; etc.").

A second and third set of questions (Goldberg Quick Big-5 and IPC-7) are meant to collect information about your personality. (Example: On a scale from 1 to 9, rating how generous you are.)

The third set of questions (Interest Profiler Form) tries to gather information about your interests. (Example: Indicating whether you strongly dislike, dislike, like, or strongly like the activity "investigate the cause of a fire".)

The remaining sets of questions ask you to make predictions about the performance of hypothetical job applicants if they were hired. (Example: Job applicant # 5 scored at the 50th percentile (an average score) on a cognitive ability (intelligence) test and is applying for a job that the applicant is expected to find boring. On a percentile scale from 0 to 100, rating how well you think the applicant will perform the job if he or she were hired.) You will receive feedback about the nature of task that you are doing in the form of displays (i.e., pictures such as pie charts and line graphs) with instructions about how to interpret them.

The academic, experience, personality, and interest measures will help us to determine if people with certain characteristics perform better or worse than people with other characteristics in making the predictions. Because a major purpose of our study is to determine if academic background is related to performance on the prediction tasks, we will be asking your permission to allow the University of Minnesota to give us access to your college grade point average and number of completed academic hours as well as access to your high school percentile rank and standardized test (e.g., ACT, SAT) scores.

Risks and Benefits of Being in the Study

The study has several risks: First, and as with any disclosure made to anyone, there always is a non-zero risk of unforeseeable breach of confidentiality, security and/or anonymity. However, this risk is minimal. **ALL INFORMATION THAT YOU PROVIDE WILL BE KEPT CONFIDENTIAL AND SECURED.** Second, you might find the prediction task to be difficult and frustrating. Just do the best that you can.

There are no direct benefits to participation in this study.

Compensation:

You will receive 4 REP points shortly after your participation in this study no matter how well you perform in it. If your predictions are, on average, more accurate than those of every other participant, then you will be paid $75 by the conclusion of this study. The participant who performs second-best will be paid $50, and the participant who performs third-best will be paid $25. Financial compensation will be awarded after all subjects have completed the study, which will occur no later than August 2010.

Confidentiality:

The records of this study will be kept private. In any sort of report we might publish, we will not include any information that will make it possible to identify a subject. Research records will be stored securely and only researchers will have access to the records. Data will be confidential, but the researcher cannot guarantee complete anonymity until linking of data to academic information is completed.

Voluntary Nature of the Study:

Participation in this study is voluntary. Your decision whether or not to participate will not affect your current or future relations with the University of Minnesota. If you decide to participate, you are free to not answer any question or withdraw at any time without affecting those relationships.

Contacts and Questions:

The researcher conducting this study is David M. Klieger, J.D. You may ask any questions you have now. If you have questions later, **you are encouraged** to contact David Klieger at The Department of Psychology, Elliott Hall, 75 East River Road, Minneapolis, MN 55455, (612) 625-3529, klie0019@umn.edu. You may contact his faculty advisor, Professor Nathan R. Kuncel at the same address, but at the following different telephone number and e-mail address: (612) 625-2818, kunce001@umn.edu.

If you have any questions or concerns regarding this study and would like to talk to someone other than the researcher(s), **you are encouraged** to contact the Research Subjects' Advocate Line, D528 Mayo, 420 Delaware St. Southeast, Minneapolis, Minnesota 55455; (612) 625-1650.

You will be given a copy of this information to keep for your records.

Statement of Consent to the Study:

I have read the above information. I have asked questions and have received answers. I consent to participate in the study.

Signature:______________________________ ______ Date: ___________________

Printed name: _________________________ University ID # ___________________

Signature of parent or guardian:______________________ Date: ___________________
(If minors are involved)

Printed name: ______________________________

Signature of Investigator:________________________ Date: ___________________

Statement of Consent to the Release of Academic Information:

I consent to the release by the University of Minnesota to the primary investigator, David M. Klieger, of my college grade point average(s), number of completed academic hours, high school rank, and standardized test scores (e.g., ACT, SAT) scores.

Signature:_____________________________________ Date: ___________________

Printed name: _________________________ University ID # ___________________

Signature of parent or guardian:______________________ Date: ___________________
(If minors are involved)

Printed name: ______________________________

Signature of Investigator:___________________________ Date: ___________________

Appendix L

Academic Information & Experience Form (Fall 2009 & Spring 2010)

Please provide the information requested below. Some examples have been provided.

Academic Information:

	Your ACT, SAT, GRE, and any other standardized test scores:	Which Test (e.g., SAT, ACT, GRE, LSAT, GMAT, MCAT, PCAT, etc.)?	Approximate Date When Test Was Taken	Any Subtest Scores
Example a:	25	ACT	June 2007	English= 23; Mathematics = 20; Reading = 22; Science = 25
Example b:	1100	SAT	May 2008	580 on SAT critical reading; 520 on SAT mathematics; essay score = 9; multiple choice writing questions score = 50
c.				
d.				
e.				
f.				
g.				

Your college GPA (cumulative): _____________________________ (if not from the University of Minnesota – Twin Cities, then please note the college / university)
Your major(s): ____________________________
Your High School Rank (e.g., top 25%, 87th out of 400 students, etc.): __________________________________

Have you ever predicted the performance of anybody before? (Think about this question very broadly.) Yes / No

If "No," then please go to question (H.), below. If "Yes," then . . .

(A.) for what outcomes were you making the predictions? (e.g., for fit in your fraternity/sorority; for performance in a job for which you would be that person's boss; for a position on your baseball team; for performance in political office; whether a person would make a good significant other, etc.)? You may name up to 4 types of outcomes in the first empty column (A) below.

(B.) for how many people did you make the predictions? Write your responses in the second empty column (B) below.

(C.) for what length of time (if any) did you make the predictions? Write your responses in the third column (C) below.

(D.) was the process of making the predictions a formal process? Write your responses in the fourth column (D) below.

(E.) did you receive any training to make the predictions? Write your responses in the fifth column (E) below.

(F.) was the training (if any) formal? Write your responses in the sixth column (F) below.

(G.) for how long did you receive training (if any)? Write your responses in the seventh column (G) below.

(H.) indicate your expectations for this study. Circle your responses in the eighth column (H) below.

	(A.) Outcomes	(B.) # People	(C.) Time Length	(D.) Formal Prediction Process? (Yes/No)	(E.) Training? (Yes/No)	(F.) Formal Training? (Yes/No)	(G.) Length of Training	(H.) Expectations for this Study
Example:	for performance as junior sales clerk in store where I was senior sales clerk	11	during a 2-yr. period	Y / N	Y / N	Y / N	1½ hours (spoke informally with other senior sales clerk)	In terms of your accuracy, how do you think that you will predict job performance in this study? (Circle one choice.): |---------|--------|------|------|--------|---------| 0 Extremely poorly; 1 Moderatel y poorly; 2 Slightly poorly; 3 So-so; 4 Slightl y well; 5 Moderatel y well; 6 Extremely well
a.				Y / N	Y / N	Y / N		In terms of how you will perform relative to other participants in this study, how do you think that you will predict job performance in this study? (Circle one choice.): |---------|--------|------|------|--------|---------| 0 Extremely below average; 1 Moderatel y below average; 2 Slightly below averag e; 3 Aver age; 4 Slightl y above avera ge; 5 Moderatel y above average; 6 Extremely above average
b.				Y / N	Y / N	Y / N		
c.				Y / N	Y / N	Y / N		
d.				Y / N	Y / N	Y / N		

(I.) Have you taken any statistics courses before? Yes / No If "yes", please describe: _______________________

(J.) Have you taken any courses in decision-making or decision theory before? Yes / No If "yes", please describe: ____________________________________

(K.) Are you male or female? (Males and females might not receive equivalent training opportunities, or some other gender-based difference might relate to the outcomes of this study.): ____________

Appendix M

Goldberg Personality Inventory (Fall 2009 & Spring 2010)

Please use this list of common human traits to describe yourself as accurately as possible. Describe yourself as you see yourself at the present time, not as you wish to be in the future. Describe yourself as you are generally or typically, compared with other persons you know of the same sex and of roughly the same age.

For each trait, please write down the number in the space provided beside the word, indicating how accurately that trait describes you, using the following rating scale.

Inaccurate					Accurate			
Extremely	Very	Quite	Slightly	Neither	Slightly	Quite	Very	Extremely
1	2	3	4	5	6	7	8	9

____	Active	____	Extraverted	____	Negligent	____	Trustful
____	Agreeable	____	Fearful	____	Nervous	____	Unadventurous
____	Anxious	____	Fretful	____	Organized	____	Uncharitable
____	Artistic	____	Generous	____	Philosophical	____	Uncooperative
____	Assertive	____	Haphazard	____	Pleasant	____	Uncreative
____	Bashful	____	Harsh	____	Practical	____	Undemanding
____	Bold	____	Helpful	____	Prompt	____	Undependable
____	Bright	____	High-Strung	____	Quiet	____	Unemotional
____	Careful	____	Imaginative	____	Relaxed	____	Unenvious
____	Careless	____	Imperceptive	____	Reserved	____	Unexcitable
____	Cold	____	Imperturbable	____	Rude	____	Unimaginative
____	Complex	____	Impractical	____	Self-Pitying	____	Uninquisitive
____	Conscientious	____	Inconsistent	____	Selfish	____	Unintellectual
____	Considerate	____	Inefficient	____	Shallow	____	Unintelligent
____	Cooperative	____	Inhibited	____	Shy	____	Unkind
____	Creative	____	Innovative	____	Simple	____	Unreflective
____	Daring	____	Insecure	____	Sloppy	____	Unrestrained
____	Deep	____	Intellectual	____	Steady	____	Unsophisticated
____	Demanding	____	Introspective	____	Sympathetic	____	Unsympathetic
____	Disorganized	____	Introverted	____	Systematic	____	Unsystematic
____	Distrustful	____	Irritable	____	Talkative	____	Untalkative
____	Efficient	____	Jealous	____	Temperamental	____	Verbal
____	Emotional	____	Kind	____	Thorough	____	Vigorous
____	Energetic	____	Moody	____	Timid	____	Warm
____	Envious	____	Neat	____	Touchy	____	Withdrawn

Appendix N

IPC-7 Personality Inventory (Fall 2009 & Spring 2010)

developed by Tellegen, Waller, & Grove

Please indicate whether you agree or disagree that the following words or phrases describe you. This scale runs from strong disagreement with the accuracy of description to strong agreement with the accuracy of description.

		STRONGLY DISAGREE	DISAGREE	AGREE	STRONGLY AGREE
1.	not easily upset	A	B	C	D
2.	don't talk much, uncommunicative	A	B	C	D
3.	conventional	A	B	C	D
4.	exceptional, special	A	B	C	D
5.	dangerous to others, harmful	A	B	C	D
6.	others think I am quarrelsome and contentious	A	B	C	D
7.	quiet	A	B	C	D
8.	reserved, distant	A	B	C	D
9.	often feel guilty for no reason	A	B	C	D
10.	important, significant	A	B	C	D
11.	easy on others, lenient	A	B	C	D
12.	rather put up a fight than make a concession	A	B	C	D
13.	average, unremarkable	A	B	C	D
14.	disgusting, horrible	A	B	C	D
15.	consistent, predictable	A	B	C	D
16.	strong, forceful	A	B	C	D
17.	lively, animated	A	B	C	D
18.	often irritated by minor setbacks	A	B	C	D
19.	tough, uncompromising	A	B	C	D

		STRONGLY DISAGREE	DISAGREE	AGREE	STRONGLY AGREE
20.	talkative	A	B	C	D
21.	like things to be a bit disorganized and chaotic	A	B	C	D
22.	prefer to be alone, a loner	A	B	C	D
23.	playful	A	B	C	D
24.	can put worries out of mind	A	B	C	D
25.	gregarious, sociable	A	B	C	D
26.	vicious, nasty	A	B	C	D
27.	do things in an orderly and systematic manner	A	B	C	D
28.	often jumpy and jittery	A	B	C	D
29.	hold traditional values and beliefs	A	B	C	D
30	awful, terrible	A	B	C	D
31.	stubborn, obstinate	A	B	C	D
32.	an ordinary, everyday person	A	B	C	D
33.	wicked, evil	A	B	C	D
34.	get into arguments, argumentative	A	B	C	D
35.	deserve to be admired	A	B	C	D
36.	peppy, spirited	A	B	C	D
37.	high ranking, powerful	A	B	C	D
38.	do not worry about the little things	A	B	C	D
39.	feelings are easily hurt	A	B	C	D
40.	keep belongings neat and tidy	A	B	C	D
41.	politically radical, hold revolutionary views	A	B	C	D
42.	prompt, punctual, get things done on time	A	B	C	D
43.	don't let many things bother or frustrate me	A	B	C	D
44.	cautious, circumspect	A	B	C	D
45.	believe that strict discipline at home would prevent most of the crime in society today	A	B	C	D
46.	odd, peculiar	A	B	C	D
47.	impressive, remarkable	A	B	C	D
48.	headstrong, willful	A	B	C	D
49.	unusual, unconventional	A	B	C	D

		STRONGLY DISAGREE	DISAGREE	AGREE	STRONGLY AGREE
50.	excellent, first-rate	A	B	C	D
51.	dislike arguments and conflict	A	B	C	D
52.	well-organized	A	B	C	D
53.	deserve to be hated	A	B	C	D
54.	like to improvise, "play things by ear"	A	B	C	D
55.	nervous, high-strung	A	B	C	D
56.	cruel, mean	A	B	C	D
57.	believe that most parents are too permissive, let their children get away with too much	A	B	C	D
58.	outstanding, superior	A	B	C	D
59.	often feel sorry for myself	A	B	C	D
60.	strange	A	B	C	D
61.	spontaneous, impulsive	A	B	C	D
62.	depraved, perverted	A	B	C	D
63.	try to avoid difficulties with other people	A	B	C	D
64.	mentally disturbed, sick	A	B	C	D
65.	not exceptional, not that special	A	B	C	D
66.	like to be with people, sociable	A	B	C	D
67.	treacherous, disloyal	A	B	C	D
68.	conservative	A	B	C	D
69.	like to have place for everything and everything in its place	A	B	C	D
70.	progressive, favor social reform	A	B	C	D

Appendix O

RIASEC Interest Profiler Form (Fall 2009 & Spring 2010)

Blue text or frame = unique to Fall 2009 study
Red text or frame = unique to Spring 2010 study

Welcome to the Interest Profiler. This inventory is designed to help you explore your vocational interests by rating the extent to which you would like to do certain activities. To complete this questionnaire, select the description that most closely represents **how you feel** about each of the activities.

		STRONGLY DISLIKE	DISLIKE	UNSURE	LIKE	STRONGLY LIKE
1.	Build kitchen cabinets	A	B	C	CD	DE
2.	Manage a retail store	A	B	C	CD	DE
3.	Study the movement of planets	A	B	C	CD	DE
4.	Help conduct a therapy group session	A	B	C	CD	DE
5.	Calculate the wages of employees	A	B	C	CD	DE
6.	Develop a spreadsheet using computer software	A	B	C	CD	DE
7.	Examine blood samples using a microscope	A	B	C	CD	DE
8.	Investigate the cause of a fire	A	B	C	CD	DE
9.	Paint sets for plays	A	B	C	CD	DE
10.	Start your own business	A	B	C	CD	DE
11.	Negotiate business contracts	A	B	C	CD	DE
12.	Proofread records or forms	A	B	C	CD	DE
13.	Inventory supplies using a hand-held computer	A	B	C	CD	DE
14.	Design sets for plays	A	B	C	CD	DE
15.	Represent a client in a lawsuit	A	B	C	CD	DE
16.	Develop a way to better predict the weather	A	B	C	CD	DE
17.	Work in a biology lab	A	B	C	CD	DE
18.	Write scripts for movies or television shows	A	B	C	CD	DE
19.	Lay brick or tile	A	B	C	CD	DE

		STRONGLY DISLIKE	DISLIKE	UNSURE	LIKE	STRONGLY LIKE
20.	Market a new line of clothing	A	B	C	CD	DE
21.	Test the quality of parts before shipment	A	B	C	CD	DE
22.	Invent a replacement for sugar	A	B	C	CD	DE
23.	Perform jazz or tap dance	A	B	C	CD	DE
24.	Take care of children at a day-care center	A	B	C	CD	DE
25.	Sell merchandise at a department store	A	B	C	CD	DE
26.	Record rent payments	A	B	C	CD	DE
27.	Repair and install locks	A	B	C	CD	DE
28.	Manage a clothing store	A	B	C	CD	DE
29.	Keep inventory records	A	B	C	CD	DE
30.	Set up and operate machines to make products	A	B	C	CD	DE
31.	Do laboratory tests to identify diseases	A	B	C	CD	DE
32.	Study weather conditions	A	B	C	CD	DE
33.	Edit movies	A	B	C	CD	DE
34.	Teach a high-school class	A	B	C	CD	DE
35.	Stamp, sort, and distribute mail for an organization	A	B	C	CD	DE
36.	Help people with personal or emotional problems	A	B	C	CD	DE
37.	Operate a beauty salon or barber shop	A	B	C	CD	DE
38.	Monitor a machine on an assembly line	A	B	C	CD	DE
39.	Repair household appliances	A	B	C	CD	DE
40.	Write books or plays	A	B	C	CD	DE
41.	Play a musical instrument	A	B	C	CD	DE
42.	Teach children how to read	A	B	C	CD	DE
43.	Load computer software into a large computer network	A	B	C	CD	DE
44.	Study ways to reduce water pollution	A	B	C	CD	DE
45.	Give career guidance to people	A	B	C	CD	DE

		STRONGLY DISLIKE	DISLIKE	UNSURE	LIKE	STRONGLY LIKE
46.	Raise fish in a fish hatchery	A	B	C	CD	DE
47.	Compose or arrange music	A	B	C	CD	DE
48.	Operate a calculator	A	B	C	CD	DE
49.	Assemble electronic parts	A	B	C	CD	DE
50.	Drive a truck to deliver packages to offices and homes	A	B	C	CD	DE
51.	Perform rehabilitation therapy	A	B	C	CD	DE
52.	Do volunteer work at a non-profit organization	A	B	C	CD	DE
53.	Teach an individual an exercise routine	A	B	C	CD	DE
54.	Conduct chemical experiments	A	B	C	CD	DE
55.	Draw pictures	A	B	C	CD	DE
56.	Buy and sell stocks and bonds	A	B	C	CD	DE
57.	Create special effects for movies	A	B	C	CD	DE
58.	Teach sign language to people with hearing disabilities	A	B	C	CD	DE
59.	Manage a department within a large company	A	B	C	CD	DE
60.	Keep shipping and receiving records	A	B	C	CD	DE

Appendix P

Questions About Confidence Level and Self-Perceptions of Strategies (Insight) for the Immediately Previous Block of Prediction (Fall 2009 & Spring 2010)

Form For Block 1 and When the Disordinal Interaction Has Not Been Explicitly Mentioned Yet and No Disordinal Interaction Feedback Has Been Provided Yet:

In terms of your accuracy, how do you think that you did on the previous 25 predictions of job performance?

|-------------------|-------------------|------------------|--------------------|-------------------|------------------
-|

0	1	2	**3**	4	5	**6**
Extremely poorly	Moderately poorly	Slightly poorly	So-so	Slightly well	Moderately well	Extremely well

In terms of how you performed relative to other participants in this study, how do you think that you did on the previous 25 predictions of job performance?

|-------------------|-------------------|------------------|--------------------|-------------------|------------------
-|

0	1	2	**3**	4	5	**6**
Extremely below average	Moderately below average	Slightly below average	Average	Slightly above average	Moderately above average	Extremely above average

Please describe your strategies that you used to make the previous 25 predictions of job performance:

What do you think the relationship is between cognitive ability test score and job performance?

In making the previous 25 predictions, did you try to use information <u>other than or in addition to</u> cognitive ability test score or how interesting or boring the job is? Yes/No

If "yes", then what information did you try to use, and how?

Form For When the Disordinal Interaction Has Been Explicitly Mentioned Already and/or When Disordinal Interaction Feedback Has Been Provided Already:

In terms of your accuracy, how do you think that you did on the previous 25 predictions of job performance?

|-------------------|-------------------|------------------|--------------------|-------------------|-------------------|

0	1	2	**3**	4	5	**6**
Extremely poorly	Moderately poorly	Slightly poorly	So-so	Slightly well	Moderately well	Extremely well

In terms of how you performed relative to other participants in this study, how do you think that you did on the previous 25 predictions of job performance?

|-------------------|-------------------|------------------|--------------------|-------------------|-------------------|

0	1	2	**3**	4	5	**6**
Extremely below average	Moderately below average	Slightly below average	Average	Slightly above average	Moderately above average	Extremely above average

Please describe your strategies that you used to make the previous 25 predictions of job performance:

Did you try to make use of the disordinal interaction information in making the previous 25 predictions of job performance? Yes/No

If "yes", then how did you try to use it?

In making the previous 25 predictions, did you try to use information <u>other than or in addition to</u> cognitive ability test score, how interesting or boring the job is, or the disordinal interaction? Yes/No

If "yes", then what information did you try to use, and how?

Appendix Q

Questions About Confidence Level and Self-Perceptions of Strategies (Insight) for Predictions Overall
(Provided to Subjects After All Predictions Have Been Made)

In terms of your accuracy, how do you think that you did on the **entire 125** predictions of job performance? Please circle one of the choices below:

|-----------------|-----------------|----------------|------------------|-----------------|------------------|

0	1	2	**3**	4	5	**6**
Extremely poorly	Moderately poorly	Slightly poorly	So-so	Slightly well	Moderately well	Extremely well

How do you think that you did on the **entire 125** predictions of job performance compared to other participants in this study? Please circle one of the choices below:

|-----------------|-----------------|----------------|------------------|-----------------|------------------|

0	1	2	**3**	4	5	**6**
Extremely below average	Moderately below average	Slightly below average	Average	Slightly above average	Moderately above average	Extremely above average

Did your strategies change over the course of making the 125 predictions? Yes or No

If your strategies did change over the course of making the 125 predictions, then when did they change? (e.g., after the third set of 25 predictions, after prediction # 82, etc.): ___________________________

How did your strategies change (if they changed at all) over the course of making the 125 predictions?

Why did you change or not change your prediction strategies over the course of making the 125 predictions?

Appendix R

TABLES

Table 1. *Lens Model Parameters*

1. OUTCOME (CRITERION), TASK (ENVIRONMENTAL), AND COGNITIVE (JUDGMENT) PARAMETERS		
Parameter	Description	Statistical Relationships
Y_e	criterion value (as actually observed/measured)	
$\hat{Y}_e$	criterion value predicted from mechanical data combination	$\hat{Y}_e = b_{1e}X_1 + \ldots + b_{ke}X_k$
Y_s	clinician's judgment; criterion value predicted from clinical data combination	
$\hat{Y}_s$	bootstrapped ("model of man") prediction; judgment model prediction; policy model prediction; predicted value from the policy model developed by regressing observed cue values on the clinician's judgments	Bootstrapped model ("model of man"): $\hat{Y}_s = b_{1s}X_1 + \ldots + b_{ks}X_k$
$X_1 \ldots X_k$	stimulus dimensions; cues; independent variables; predictors; indicators	Mechanical data combination: $\hat{Y}_e = b_{1e}X_1 + \ldots + b_{ke}X_k$ Bootstrapped model ("model of man"): $\hat{Y}_s = b_{1s}X_1 + \ldots + b_{ks}X_k$
$r_{1e} \ldots r_{ke}$	ecological validities; validities that can be mechanically derived from the mechanical data combination (possibly used as weights for the mechanical data combination)	
$r_{1s} \ldots r_{ks}$	utilization coefficients; validities that can be mechanically derived from the clinical data combination to describe each perceived cue-criterion relationship independent of the other cue-criterion relationships	
r_{ij}	cue redundancy; cue inter-correlation	
$b_{1e} \ldots b_{ke}$	ecological (environmental) weights used in mechanical data combination (sometimes describing the relative importance of each cue-criterion relationship in the environment)	$\hat{Y}_e = b_{1e}X_1 + \ldots + b_{ke}X_k$
$b_{1s} \ldots b_{ks}$	judgment model weights; bootstrapped weights; policy model weights; weights that can be mechanically derived from the clinical data combination (and that some use to describe the relative importance of each cue-criterion relationship in the judgment)	$\hat{Y}_s = b_{1s}X_1 + \ldots + b_{ks}X_k$

Table 1. – *cont'd.*

2. **FUNCTIONAL VALIDITY PARAMETERS: r_a AND THE LINEAR COMPONENT**		
Parameter	Description	Statistical Relationships
r_a	criterion-related validity of the clinical judgment; achievement index	$r_a = r_{Ye,Ys} = GR_eR_s + C\sqrt{(1-R_e^2)}\sqrt{(1-R_s^2)}$ where GR_eR_s = the mechanical component, and $C\sqrt{(1-R_e^2)}\sqrt{(1-R_s^2)}$ = the unmodeled component (traditionally known by the overly restrictive description "the configural component") When $C\sqrt{(1-R_e^2)}\sqrt{(1-R_s^2)} \approx 0$ (which many studies arrange or assume), $r_a = r_{Ye,Ys} \approx GR_eR_s$
R_e	criterion-related validity of mechanical data combination; environmental validity or predictability	$R_e = r_{Ye,\hat{Y}e}$
R_s	cognitive control (Hammond & Summers, 1972); traditionally known by the misnomers "response linearity" and "cognitive consistency" R_c (Cooksey, 1996); criterion-related validity of a bootstrapped model of the judge's prediction strategies	$R_s = r_{Ys,\hat{Y}s} \le R_c$ where $(1-R_s{}^2)$ indexes variation in profile judgments in general (variation around $\hat{Y}_j$, where j = the jth profile), and $(1-R_c{}^2)$ indexes variation in repeated judgments of a single profile (variation around $\overline{Y}_j$) (Cooksey, 1996)
G (sometimes r_m)	mechanical knowledge; matching index; sometimes narrowly referred to as linear knowledge (although knowledge could be mechanical and non-linear); does *not* necessarily reflect extent to which ecological b_e weights match judgment b_s weights (Karelaia & Hogarth, 2008)	$G = r_{\hat{Y}e,\hat{Y}s}$
GR_e (sometimes r_m)	validity of the bootstrapped model ("model of man") for predicting the actual criterion Y_e	$GR_e = r_{Ye,\hat{Y}s}$

Table 1. – *cont'd.*

3. COGNITIVE (JUDGMENT) RESIDUALS & 4. FUNCTIONAL VALIDITY PARAMETERS: THE UNMODELED COMPONENT		
Parameter	Description	Statistical Relationships
$1-R_c^{\ 2}$	cognitive inconsistency; variability due to random error in bootstrapped model; test-retest (temporal stability) unreliability (1 – r_{jj}, where j = the jth profile)	$1-R_c^{\ 2}$ = *var(random error)* = $1 - r_{jj}$; see Cooksey (1996) pp. 207-208 for alternative formulation; $1-R_c^{\ 2}$
z_s	residual of bootstrapped model	$z_s = Y_s - \hat{Y}_s$
r_z	validity of residual of bootstrapped model for predicting the criterion value (as actually measured, Y_e)	$r_z = C\sqrt{\left(1-R_e^2\right)}$ (Camerer, 1981a; Einhorn, 1974)
$1-R_s^{\ 2}$	residual variance (of the bootstrapped model); error variance (of the bootstrapped model); lack of cognitive control; (1) variability due to absence of cues in bootstrapped model + (2) variability due to absence of nonlinear cue functional forms in bootstrapped model + (3) variability due to absence of cue interactions in bootstrapped model + (4) variability due to random error in bootstrapped model (Camerer, 1981a; Einhorn, 1974)	$1-R_s^{\ 2} = \mathrm{var}(Y_s - \hat{Y}_s) = \mathrm{var}(z_s) =$ *var(absent cues)* + *var(absent cue nonlinearity)* + *var(absent cue interactions)* + *var(random error)*
C	unmodeled knowledge; unmodeled agreement; residual correlation; clinician's reliance on cues ("broken legs") and cue functional forms absent from the mechanical model + chance agreement between mechanical and expert's errors (Cooksey, 1996); traditionally known by overly restrictive description "configural cue use" (Cooksey, 1996)	$C = r_{Ye-\hat{Y}e, Ys-\hat{Y}s}$
$C\sqrt{\left(1-R_e^{\ 2}\right)}$	validity of the residual of the clinical data combination used to index the extent to which the validity of clinical data combination can be improved (Camerer, 1981a; Einhorn, 1974)	$C\sqrt{\left(1-R_e^{\ 2}\right)} = r_z = r_{Ye, Ys-\hat{Y}s}$

Table 2. *Examples of Lens Model Parameters in Naturalistic Environments*

	Sample	Criterion	Cues/Predictors	# Clinicians	Mean # Clinical Judgments Per Clinician	SD for # Clinical Judgments Per Clinician	Mean ra	Re (# pred)	Mean G	Unmodeled ra Component	Mean rz	Mean C	Mean Rs
1	Workers in Home Office and First-Line & Middle Management; No Executives	Fit	7: (1) cognitive ability composite, (2) achievement motivation, (3) responsibility/socialization, (4) leadership/dominance, (5) role play task, (6) behavioral interview, (7) assessment center tests	13	13	10.1	0.13	0.22 (161)	0.52	0.06	0.16	0.16	0.89
2	First-Line and Middle Management	Job Performance	5: (1) cognitive ability composite, (2) achievement motivation, (3) responsibility/socialization, (4) leadership/dominance, (5) in-basket task, behavioral interview, role play task composite	7	11.1	20.7	0.09	0.29 (154)	0.76	-0.08	-0.05	-0.07	0.70
3	Management & Possibly a Very Small # of Executives	Overall Job Performance Rating By Boss	7: (1) cognitive ability composite, (2) achievement motivation, (3) responsibility/socialization, (4) leadership/dominance, (5) in-basket task, (6) behavioral interview, (7) assessment center tests	14	25.5	20.3	-0.01	0.25 (183)	0.38	-0.12	-0.27	-0.26	0.88

Notes. Mean ra = mean validity of clinical data combination across clinicians; Re (# pred) = validity of mechanical data combination after Wherry fixed effects shrinkage to the population level (number of predictions on which shrinkage is based); Mean G = mean linear knowledge; Unmodeled component = $C\sqrt{(1-R_e^2)}\sqrt{(1-R_s^2)}$ from standard Lens Model Equation; Mean rz = mean validity of residual from judges' policies ($C\sqrt{(1-R_e^2)}$); Mean C = mean unmodeled knowledge across judges; Mean Rs = mean cognitive control across judges. All means were weighted based on the number of judgments for each clinician.

Table 3. *Importance of the Disordinal Interaction: Sample Level (Fall 2009 & Spring 2010)*

	Number of Cases (Profiles)	Disordinal Interaction Mechanical Model R^2 (and R)	Purely Linear Mechanical Model R^2 (and R)	ΔR^2 from Including the Disordinal Interaction	Relative Weight C_{xj} of Disordinal Interaction Term
Fall 2009 Dataset	100 (4 blocks of 25)	0.171 (0.414)	0.018 (0.134)	0.153 (850% increase)	0.0852
	25 (Block 1)	0.074 (0.273)	0.021 (0.146)	0.053 (252% increase)	0.0637
	25 (Block 2)	0.269 (0.519)	0.012 (0.110)	0.257 (2141% increase)	0.2631
	25 (Block 3)	0.261 (0.511)	0.038 (0.195)	0.223 (587% increase)	0.2422
	25 (Block 4)	0.214 (0.463)	0.023 (0.151)	0.191 (830% increase)	0.2025
Spring 2010 Dataset	125 (5 blocks of 25)	0.518 (0.720)	0.122 (0.350)	0.396 (325% increase)	0.2452
	25 (Block 1)	0.733 (0.856)	0.551 (0.742)	0.182 (33% increase)	0.4625
	25 (Block 2)	0.648 (0.805)	0.177 (0.421)	0.471 (266% increase)	0.3478
	25 (Block 3)	0.464 (0.681)	0.002 (0.049)	0.462 (23,100% increase)	0.4712
	25 (Block 4)	0.561 (0.749)	0.046 (0.215)	0.515 (1120% increase)	0.5358
	25 (Block 5)	0.487 (0.698)	0.252 (0.502)	0.235 (93% increase)	0.5597

Note. For both datasets, the blocks do not appear in the same order for all subjects, even within the same treatment (feedback) condition for the same dataset. Furthermore, the nature and quantity of feedback about the disordinal interaction varied across treatment conditions within each of the datasets. For the Spring 2010 dataset, blocks were randomly chosen from a population of data that reflected a disordinal interaction. For the Fall 2009 dataset only, the blocks were selectively (not randomly) chosen from their respective population based upon how they would appear in scatterplots where the relationship between cognitive ability and performance was depicted separately for (a) jobs above average in how interesting they were and (b) jobs below average in how interesting they were. Although the Fall 2009 0-6 interesting/boring scale consisted of an odd number of possible scores, this dichotomization for the scatterplots was based upon those scores prior to their being transformed to the 0-6 scale (i.e., when they were dichotomously scored 0 or 1 prior to the addition of random variability).

Table 4. *Importance of the Disordinal Interaction: Population Level (Fall 2009 & Spring 2010)*

	Number of Cases (Profiles) Used in Simulation of the Disordinal Interaction (Population Size)	Disordinal Interaction Mechanical Model R^2 (and R)	Purely Linear Mechanical Model R^2 (and R)	ΔR^2 from Including the Disordinal Interaction	Relative Weight C_{xj} of Disordinal Interaction Term
Fall 2009 Data	2,000	0.121 (0.348)	0.028 (0.168)	0.093 (332% increase)	0.0595
Spring 2010 Data	10,000	0.490 (0.700)	0.134 (0.367)	0.356 (266% increase)	0.2208

Table 5. *Variables (Fall 2009 & Spring 2010)*

Variable Name	Variable Description	Dependent or Independent Variable (DV or IV)	Level (1st = Within Subjects; 2nd = Between Subjects) [73]	Additional Scale Information (or N/A If Variable Was Not Analyzed)	
				Fall 2009 Data (4 blocks; 100 predictions)	**Spring 2010 Data (5 blocks; 125 predictions)**
IDENTIFICATION VARIABLES					
SUBNUM	subject identification number	IV	2nd		
Form	form number (based on Ordering and Feedback)		2nd	2 sets of 3 ordinal values	120 sets of 3 ordinal values
Ordering	item and feedback ordering (to control for ordering effects)		2nd	2 nominal values (2 orderings); limited randomized blocks design	120 nominal values (5! permutations of 5 blocks); fully randomized blocks design
Feedback	experimental feedback treatment (group)		2nd	3 ordinal values (3 groups, each receiving different amounts of feedback information)	
D1	dummy coded variable to indicate feedback treatment (group)	IV	2nd	D1D2 coding is 00 for treatment 1, 01 for treatment 2, and 10 for treatment 3	
D2	dummy coded variable to indicate feedback treatment (group)	IV	2nd		

[73] All 1st level (within-subjects variables) for a multilevel analysis can act as 2nd-level (between-subjects variables) for another context. For example, they act as 2nd-level variables in the intercorrelation matrices in Table 10 and Table 19.

Table 5 – *cont'd.*

Variable Name	Variable Description	Dependent or Independent Variable (DV or IV)	Level (1st = Within Subjects; 2nd = Between Subjects) [74]	Additional Scale Information (or N/A If Variable Was Not Analyzed)	
				Fall 2009 Data (4 blocks; 100 predictions)	Spring 2010 Data (5 blocks; 125 predictions)
VARIABLES MEASURING JUDGMENT ACCURACY (ra, SS)					
• *Judgment (Clinical) Criterion-Related Validity (ra):*					
ra	correlation between clinical predictions and observed outcomes for all predictions (100 or 125)	DV & IV	2nd	Pearson Product Moment correlation	
ra1	correlation between clinical predictions and observed outcomes for 1st block of 25 predictions (prior to the provision of any feedback)	DV & IV	1st & 2nd	Pearson Product Moment correlation	
ra2	correlation between clinical predictions and observed outcomes for 2nd block of 25 predictions	DV & IV	1st & 2nd	Pearson Product Moment correlation	
ra3	correlation between clinical predictions and observed outcomes for 3rd block of 25 predictions	DV & IV	1st & 2nd	Pearson Product Moment correlation	
ra4	correlation between clinical predictions and observed outcomes for 4th block of 25 predictions	DV & IV	1st & 2nd	Pearson Product Moment correlation	
ra5	correlation between clinical predictions and observed outcomes for 5th block of 25 predictions	DV & IV	1st & 2nd	N/A	Pearson Product Moment correlation
Fisherra	Fisher r-to-z transformed ra (to normalize the variable)	DV & IV	2nd	normally distributed Fisher z	
Fisherra1	Fisher r-to-z transformed ra1 (to normalize the variable)	DV & IV	1st & 2nd	normally distributed Fisher z	
Fisherra2	Fisher r-to-z transformed ra2 (to normalize the variable)	DV & IV	1st & 2nd	normally distributed Fisher z	
Fisherra3	Fisher r-to-z transformed ra3 (to normalize the variable)	DV & IV	1st & 2nd	normally distributed Fisher z	
Fisherra4	Fisher r-to-z transformed ra4 (to normalize the variable)	DV & IV	1st & 2nd	normally distributed Fisher z	
Fisherra5	Fisher r-to-z transformed ra5 (to normalize the variable)	DV & IV	1st & 2nd	N/A	normally distributed Fisher z

[74] All 1st level (within-subjects variables) for a multilevel analysis can act as 2nd-level (between-subjects variables) for another context. For example, they act as 2nd-level variables in the intercorrelation matrix in Table 10 and Table 19.

Table 5 – *cont'd.*

Variable Name	Variable Description	Dependent or Independent Variable (DV or IV)	Level (1st = Within Subjects; 2nd = Between Subjects) [75]	Additional Scale Information (or N/A If Variable Was Not Analyzed): Fall 2009 Data (4 blocks; 100 predictions)	Additional Scale Information (or N/A If Variable Was Not Analyzed): Spring 2010 Data (5 blocks; 125 predictions)
• *Skill Score (SS):*					
SS	skill score for all predictions	DV & IV	2nd	value bounded by -∞ and 1 (1 = perfect prediction)	
SS1	skill score for 1st block of 25 predictions (prior to the provision of any feedback)	DV & IV	1st & 2nd	value bounded by -∞ and 1 (1 = perfect prediction)	
SS2	skill score for 2nd block of 25 predictions	DV & IV	1st & 2nd	value bounded by -∞ and 1 (1 = perfect prediction)	
SS3	skill score for 3rd block of 25 predictions	DV & IV	1st & 2nd	value bounded by -∞ and 1 (1 = perfect prediction)	
SS4	skill score for 4th block of 25 predictions	DV & IV	1st & 2nd	value bounded by -∞ and 1 (1 = perfect prediction)	
SS5	skill score for 5th block of 25 predictions	DV & IV	1st & 2nd	N/A	value bounded by -∞ and 1 (1 = perfect prediction)
LENS MODEL (ra) COMPONENTS (Rs, G, Cstat, and rz)					
• *Cognitive Control (Rs):*					
Rslin	Rs for linear bootstrapped model (without the disordinal interaction) for all predictions (100 or 125)	DV & IV	2nd	multiple correlation value	
Rslin1	Rs for linear bootstrapped model (without the disordinal interaction) for 1st block of 25 predictions (prior to the provision of any feedback)	DV & IV	1st & 2nd	multiple correlation value	
Rslin2	Rs for linear bootstrapped model (without the disordinal interaction) for 2nd block of 25 predictions	DV & IV	1st & 2nd	multiple correlation value	
Rslin3	Rs for linear bootstrapped model (without the disordinal interaction) for 3rd block of 25 predictions	DV & IV	1st & 2nd	multiple correlation value	
Rslin4	Rs for linear bootstrapped model (without the disordinal interaction) for 4th block of 25 predictions	DV & IV	1st & 2nd	multiple correlation value	
Rslin5	Rs for linear bootstrapped model (without the disordinal interaction) for 5th block of 25 predictions	DV & IV	1st & 2nd	N/A	multiple correlation value

[75] All 1st level (within-subjects variables) for a multilevel analysis can act as 2nd-level (between-subjects variables) for another context. For example, they act as 2nd-level variables in the intercorrelation matrix in Table 10 and Table 19.

Table 5 – *cont'd.*

Variable Name	Variable Description	Dependent or Independent Variable (DV or IV)	Level (1st = Within Subjects; 2nd = Between Subjects) [76]	Additional Scale Information (or N/A If Variable Was Not Analyzed)	
				Fall 2009 Data (4 blocks; 100 predictions)	Spring 2010 Data (5 blocks; 125 predictions)
FisherRslin	Fisher r-to-z transformed Rslin (to normalize the variable)	DV & IV	2nd	normally distributed Fisher z	
FisherRslin1	Fisher r-to-z transformed Rslin1 (to normalize the variable)	DV & IV	1st & 2nd	normally distributed Fisher z	
FisherRslin2	Fisher r-to-z transformed Rslin2 (to normalize the variable)	DV & IV	1st & 2nd	normally distributed Fisher z	
FisherRslin3	Fisher r-to-z transformed Rslin3 (to normalize the variable)	DV & IV	1st & 2nd	normally distributed Fisher z	
FisherRslin4	Fisher r-to-z transformed Rslin4 (to normalize the variable)	DV & IV	1st & 2nd	normally distributed Fisher z	
FisherRslin5	Fisher r-to-z transformed Rslin5 (to normalize the variable)	DV & IV	1st & 2nd	N/A	normally distributed Fisher z
• *Mechanical Knowledge (G):*					
G	Knowledge (use) of the linear and purely mechanical approach for all predictions (100 or 125)	DV & IV	2nd	Pearson Product Moment correlation	
G1	Knowledge (use) of the linear and purely mechanical approach for 1st block of 25 predictions (prior to the provision of any feedback)	DV & IV	1st & 2nd	Pearson Product Moment correlation	
G2	Knowledge (use) of the linear and purely mechanical approach for 2nd block of 25 predictions	DV & IV	1st & 2nd	Pearson Product Moment correlation	
G3	Knowledge (use) of the linear and purely mechanical approach for 3rd block of 25 predictions	DV & IV	1st & 2nd	Pearson Product Moment correlation	
G4	Knowledge (use) of the linear and purely mechanical approach for 4th block of 25 predictions	DV & IV	1st & 2nd	Pearson Product Moment correlation	
G5	Knowledge (use) of the linear and purely mechanical approach for 5th block of 25 predictions	DV & IV	1st & 2nd	N/A	Pearson Product Moment correlation

[76] All 1st level (within-subjects variables) for a multilevel analysis can act as 2nd-level (between-subjects variables) for another context. For example, they act as 2nd-level variables in the intercorrelation matrix in Table 10 and Table 19.

Table 5 – *cont'd.*

Variable Name	Variable Description	Dependent or Independent Variable (DV or IV)	Level (1st = Within Subjects; 2nd = Between Subjects) [77]	Additional Scale Information (or N/A If Variable Was Not Analyzed)	
				Fall 2009 Data (4 blocks; 100 predictions)	Spring 2010 Data (5 blocks; 125 predictions)
FisherG	Fisher r-to-z transformed G (to normalize the variable)	DV & IV	2nd	normally distributed Fisher z	
FisherG1	Fisher r-to-z transformed G1 (to normalize the variable)	DV & IV	1st & 2nd	normally distributed Fisher z	
FisherG2	Fisher r-to-z transformed G2 (to normalize the variable)	DV & IV	1st & 2nd	normally distributed Fisher z	
FisherG3	Fisher r-to-z transformed G3 (to normalize the variable)	DV & IV	1st & 2nd	normally distributed Fisher z	
FisherG4	Fisher r-to-z transformed G4 (to normalize the variable)	DV & IV	1st & 2nd	normally distributed Fisher z	
FisherG5	Fisher r-to-z transformed G5 (to normalize the variable)	DV & IV	1st & 2nd	N/A	normally distributed Fisher z
• *Unmodeled Knowledge (Cstat; use of new functional form(s) and "broken leg(s)" + correlated error):*					
Cstat	Unmodeled knowledge for all predictions (100 or 125)	DV & IV	2nd	Pearson Product Moment correlation	
Cstat1	Unmodeled knowledge for 1st block of 25 predictions (prior to the provision of any feedback)	DV & IV	1st & 2nd	Pearson Product Moment correlation	
Cstat2	Unmodeled knowledge for 2nd block of 25 predictions	DV & IV	1st & 2nd	Pearson Product Moment correlation	
Cstat3	Unmodeled knowledge for 3rd block of 25 predictions	DV & IV	1st & 2nd	Pearson Product Moment correlation	
Cstat4	Unmodeled knowledge for 4th block of 25 predictions	DV & IV	1st & 2nd	Pearson Product Moment correlation	
Cstat5	Unmodeled knowledge for 5th block of 25 predictions	DV & IV	1st & 2nd	N/A	Pearson Product Moment correlation

[77] All 1st level (within-subjects variables) for a multilevel analysis can act as 2nd-level (between-subjects variables) for another context. For example, they act as 2nd-level variables in the intercorrelation matrix in Table 10 and Table 19.

Table 5 – *cont'd.*

Variable Name	Variable Description	Dependent or Independent Variable (DV or IV)	Level (1st = Within Subjects; 2nd = Between Subjects) [78]	Additional Scale Information (or N/A If Variable Was Not Analyzed)	
				Fall 2009 Data (4 blocks; 100 predictions)	Spring 2010 Data (5 blocks; 125 predictions)
FisherCstat	Fisher r-to-z transformed Cstat (to normalize the variable)	DV & IV	2nd	normally distributed Fisher z	
FisherCstat1	Fisher r-to-z transformed Cstat1 (to normalize the variable)	DV & IV	1st & 2nd	normally distributed Fisher z	
FisherCstat2	Fisher r-to-z transformed Cstat2 (to normalize the variable)	DV & IV	1st & 2nd	normally distributed Fisher z	
FisherCstat3	Fisher r-to-z transformed Cstat3 (to normalize the variable)	DV & IV	1st & 2nd	normally distributed Fisher z	
FisherCstat4	Fisher r-to-z transformed Cstat4 (to normalize the variable)	DV & IV	1st & 2nd	normally distributed Fisher z	
FisherCstat5	Fisher r-to-z transformed Cstat5 (to normalize the variable)	DV & IV	1st & 2nd	N/A	normally distributed Fisher z
• *Criterion-Related Validity of Unmodeled Knowledge (rz):*					
rz	Criterion-related validity of unmodeled knowledge for all predictions (100 or 125)	DV & IV	2nd	Pearson Product Moment correlation	
rz1	Criterion-related validity of unmodeled knowledge for 1st block of 25 predictions (prior to the provision of any feedback)	DV & IV	1st & 2nd	Pearson Product Moment correlation	
rz2	Criterion-related validity of unmodeled knowledge for 2nd block of 25 predictions	DV & IV	1st & 2nd	Pearson Product Moment correlation	
rz3	Criterion-related validity of unmodeled knowledge for 3rd block of 25 predictions	DV & IV	1st & 2nd	Pearson Product Moment correlation	
rz4	Criterion-related validity of unmodeled knowledge for 4th block of 25 predictions	DV & IV	1st & 2nd	Pearson Product Moment correlation	
rz5	Criterion-related validity of unmodeled knowledge for 5th block of 25 predictions	DV & IV	1st & 2nd	N/A	Pearson Product Moment correlation

[78] All 1st level (within-subjects variables) for a multilevel analysis can act as 2nd-level (between-subjects variables) for another context. For example, they act as 2nd-level variables in the intercorrelation matrix in Table 10 and Table 19.

Table 5 – *cont'd.*

Variable Name	Variable Description	Dependent or Independent Variable (DV or IV)	Level (1st = Within Subjects; 2nd = Between Subjects) [79]	Additional Scale Information (or N/A If Variable Was Not Analyzed)	
				Fall 2009 Data (4 blocks; 100 predictions)	Spring 2010 Data (5 blocks; 125 predictions)
Fisherrz	Fisher r-to-z transformed rz (to normalize the variable)	DV & IV	2nd	normally distributed Fisher z	
Fisherrz1	Fisher r-to-z transformed rz1 (to normalize the variable)	DV & IV	1st & 2nd	normally distributed Fisher z	
Fisherrz2	Fisher r-to-z transformed rz2 (to normalize the variable)	DV & IV	1st & 2nd	normally distributed Fisher z	
Fisherrz3	Fisher r-to-z transformed rz3 (to normalize the variable)	DV & IV	1st & 2nd	normally distributed Fisher z	
Fisherrz4	Fisher r-to-z transformed rz4 (to normalize the variable)	DV & IV	1st & 2nd	normally distributed Fisher z	
Fisherrz5	Fisher r-to-z transformed rz5 (to normalize the variable)	DV & IV	1st & 2nd	N/A	normally distributed Fisher z
RELATIVE WEIGHT (IMPORTANCE) GIVEN TO THE DISORDINAL INTERACTION					
Cwtx1x2_s or C_{xj}	Relative weight (importance) for all predictions (100 or 125)	DV & IV	2nd	incremental coefficient of determination	
Cwtx1x2_1_s	Relative weight (importance) for 1st block of 25 predictions (prior to the provision of any feedback)	DV & IV	1st & 2nd	incremental coefficient of determination	
Cwtx1x2_2_s	Relative weight (importance) for 2nd block of 25 predictions	DV & IV	1st & 2nd	incremental coefficient of determination	
Cwtx1x2_3_s	Relative weight (importance) for 3rd block of 25 predictions	DV & IV	1st & 2nd	incremental coefficient of determination	
Cwtx1x2_4_s	Relative weight (importance) for 4th block of 25 predictions	DV & IV	1st & 2nd	incremental coefficient of determination	
Cwtx1x2_5_s	Relative weight (importance) for 5th block of 25 predictions	DV & IV	1st & 2nd	N/A	incremental coefficient of determination

[79] All 1st level (within-subjects variables) for a multilevel analysis can act as 2nd-level (between-subjects variables) for another context. For example, they act as 2nd-level variables in the intercorrelation matrix in Table 10 and Table 19.

Table 5 – *cont'd.*

Variable Name	Variable Description	Dependent or Independent Variable (DV or IV)	Level (1st = Within Subjects; 2nd = Between Subjects) [80]	Additional Scale Information (or N/A If Variable Was Not Analyzed)	
				Fall 2009 Data (4 blocks; 100 predictions)	Spring 2010 Data (5 blocks; 125 predictions)
CONFIDENCE IN JUDGMENT STRATEGY (SELF-REPORTED)					
• *Absolute Confidence:*					
ConfAbsOverallPre	Absolute confidence for all predictions (100 or 125) prior to making any predictions (prior to the provision of any feedback)	DV & IV	2nd	0-6 Likert scale (6 = extreme confidence)	
ConfAbs1	Absolute confidence for 1st block of 25 predictions (prior to the provision of any feedback)	DV & IV	1st & 2nd	0-6 Likert scale (6 = extreme confidence)	
ConfAbs2	Absolute confidence for 2nd block of 25 predictions	DV & IV	1st & 2nd	0-6 Likert scale (6 = extreme confidence)	
ConfAbs3	Absolute confidence for 3rd block of 25 predictions	DV & IV	1st & 2nd	0-6 Likert scale (6 = extreme confidence)	
ConfAbs4	Absolute confidence for 4th block of 25 predictions	DV & IV	1st & 2nd	0-6 Likert scale (6 = extreme confidence)	
ConfAbs5	Absolute confidence for 5th block of 25 predictions	DV & IV	1st & 2nd	N/A	0-6 Likert scale (6 = extreme confidence)
ConfAbsOverallPost	Absolute confidence for all predictions (100 or 125) after making all predictions	DV & IV	2nd	0-6 Likert scale (6 = extreme confidence)	
• *Relative Confidence (Confidence in Judgment Accuracy Relative to Accuracy of Other Subjects):*					
ConfRelOverallPre	Relative confidence for all predictions (100 or 125) prior to making any predictions (prior to the provision of any feedback)	DV & IV	2nd	0-6 Likert scale (6 = extreme confidence)	
ConfRel1	Relative confidence for 1st block of 25 predictions (prior to the provision of any feedback)	DV & IV	1st & 2nd	0-6 Likert scale (6 = extreme confidence)	
ConfRel2	Relative confidence for 2nd block of 25 predictions	DV & IV	1st & 2nd	0-6 Likert scale (6 = extreme confidence)	
ConfRel3	Relative confidence for 3rd block of 25 predictions	DV & IV	1st & 2nd	0-6 Likert scale (6 = extreme confidence)	
ConfRel4	Relative confidence for 4th block of 25 predictions	DV & IV	1st & 2nd	0-6 Likert scale (6 = extreme confidence)	
ConfRel5	Relative confidence for 5th block of 25 predictions	DV & IV	1st & 2nd	N/A	0-6 Likert scale (6 = extreme confidence)
ConfRelOverallPost	Relative confidence for all predictions (100 or 125) after making all predictions	DV & IV	2nd	0-6 Likert scale (6 = extreme confidence)	

[80] All 1st level (within-subjects variables) for a multilevel analysis can act as 2nd-level (between-subjects variables) for another context. For example, they act as 2nd-level variables in the intercorrelation matrix in Table 10 and Table 19.

Table 5 – *cont'd.*

Variable Name	Variable Description	Dependent or Independent Variable (DV or IV)	Level (1st = Within Subjects; 2nd = Between Subjects) [81]	Additional Scale Information (or N/A If Variable Was Not Analyzed)	
				Fall 2009 Data (4 blocks; 100 predictions)	Spring 2010 Data (5 blocks; 125 predictions)
INDIVIDUAL DIFFERENCES					
• *Cognitive Ability & Achievement:*					
ACT_COMP_SCR	ACT composite score	IV	2nd	score on a normally distributed scale of 1-36 (36 = highest score)	
ACT_ENGL_SCR	ACT English score	IV	2nd	score on a normally distributed scale of 1-36 (36 = highest score)	
ACT_ENGWR_SCR	ACT English Writing score	IV	2nd	score on a normally distributed scale of 1-36 (36 = highest score)	
ACT_MATH_SCR	ACT Mathematics score	IV	2nd	score on a normally distributed scale of 1-36 (36 = highest score)	
ACT_READ_SCR	ACT Reading score	IV	2nd	score on a normally distributed scale of 1-36 (36 = highest score)	
ACT_SCIRE_SCR	ACT Science Reasoning score	IV	2nd	score on a normally distributed scale of 1-36 (36 = highest score)	
CUM_GPA	Cumulative college grade point average	IV	2nd	4 = A, 3 = B, 2 = C, 1 = D, 0 = F	
TOT_ACAD_HOURS	Total number of college academic hours earned	IV	2nd	Based on University of Minnesota course credit system	
HS_RANK_PCT	High school rank	IV	2nd	Percentile	
• *Gender:*	Gender	IV	2nd	0 = Female, 1 = Male	

[81] All 1st level (within-subjects variables) for a multilevel analysis can act as 2nd-level (between-subjects variables) for another context. For example, they act as 2nd-level variables in the intercorrelation matrix in Table 10 and Table 19.

Table 5 – *cont'd.*

Variable Name	Variable Description	Dependent or Independent Variable (DV or IV)	Level (1st = Within Subjects; 2nd = Between Subjects) [82]	Additional Scale Information (or N/A If Variable Was Not Analyzed)	
				Fall 2009 Data (4 blocks; 100 predictions)	Spring 2010 Data (5 blocks; 125 predictions)
• *Personality:*					
➢ Goldberg (1990) Dimensions (Based on the Five Factor Model):					
extroGoldberg	Extroversion	IV	2nd	score on 20-180 scale (20 items scored 1-9; 180 = highest on trait)	
neurotGoldberg	Neuroticism	IV	2nd	score on 20-180 scale (20 items scored 1-9; 180 = highest on trait)	
intelGoldberg	Intellectance (Openness)	IV	2nd	score on 20-180 scale (20 items scored 1-9; 180 = highest on trait)	
agreeGoldberg	Agreeableness	IV	2nd	score on 20-180 scale (20 items scored 1-9; 180 = highest on trait)	
conGoldberg	Conscientiousness	IV	2nd	score on 20-180 scale (20 items scored 1-9; 180 = highest on trait)	
➢ Tellegen, Waller, & Grove (1987) IPC-7 Dimensions:					
pvIPC7	Positive valence (describing oneself as exceptional, important)	IV	2nd	score on 10-40 scale (10 items scored 1-4; 40 = highest on trait)	
nvIPC7	Negative valence (describing oneself as evil, immoral)	IV	2nd	score on 10-40 scale (10 items scored 1-4; 40 = highest on trait)	
pemIPC7	Positive emotionality (Extroversion)	IV	2nd	score on 10-40 scale (10 items scored 1-4; 40 = highest on trait)	
nemIPC7	Negative emotionality (Neuroticism)	IV	2nd	score on 10-40 scale (10 items scored 1-4; 40 = highest on trait)	
conIPC7	Conscientiousness	IV	2nd	score on 10-40 scale (10 items scored 1-4; 40 = highest on trait)	
agreeIPC7	Agreeableness	IV	2nd	score on 10-40 scale (10 items scored 1-4; 40 = highest on trait)	
cnvIPC7	Conventionality (Lack of Openness)	IV	2nd	score on 10-40 scale (10 items scored 1-4; 40 = highest on trait)	

[82] All 1st level (within-subjects variables) for a multilevel analysis can act as 2nd-level (between-subjects variables) for another context. For example, they act as 2nd-level variables in the intercorrelation matrix in Table 10 and Table 19.

Table 5 – *cont'd.*

Variable Name	Variable Description	Dependent or Independent Variable (DV or IV)	Level (1st = Within Subjects; 2nd = Between Subjects) [83]	Additional Scale Information (or N/A If Variable Was Not Analyzed)	
				Fall 2009 Data (4 blocks; 100 predictions)	Spring 2010 Data (5 blocks; 125 predictions)
• *RIASEC Interests:*					
ReaRIASEC	Realistic (mechanical, physical, practical)	IV	2nd	score on 10-50 scale (10 items scored 1-5; 50 = highest on trait)	
InvRIASEC	Investigative (learning, analysis, research)	IV	2nd	score on 10-50 scale (10 items scored 1-5; 50 = highest on trait)	
ArtRIASEC	Artistic (creativity, self-expression)	IV	2nd	score on 10-50 scale (10 items scored 1-5; 50 = highest on trait)	
SocRIASEC	Social (helping, teaching, team-orientation)	IV	2nd	score on 10-50 scale (10 items scored 1-5; 50 = highest on trait)	
EntRIASEC	Enterprising (persuasion, power, status)	IV	2nd	score on 10-50 scale (10 items scored 1-5; 50 = highest on trait)	
ConRIASEC	Conventional (data, numbers)	IV	2nd	score on 10-50 scale (10 items scored 1-5; 50 = highest on trait)	

[83] All 1st level (within-subjects variables) for a multilevel analysis can act as 2nd-level (between-subjects variables) for another context. For example, they act as 2nd-level variables in the intercorrelation matrix in Table 10 and Table 19.

Table 5 – *cont'd.*

Variable Name	Variable Description	Dependent or Independent Variable (DV or IV)	Level (1st = Within Subjects; 2nd = Between Subjects) [84]	Additional Scale Information (or N/A If Variable Was Not Analyzed)	
				Fall 2009 Data (4 blocks; 100 predictions)	Spring 2010 Data (5 blocks; 125 predictions)
• *Experience/Expertise:*					
numpredout	Number of different criteria (dependent variables) for which predictions of human behavior have been made prior to study participation	IV	2nd	+1 for each	
numratees	Number of different criteria (dependent variables) for which predictions have been made prior to study participation	IV	2nd	+1 for each	
lengthtimepred	Aggregate length of time over which predictions have been made prior to study participation	IV	2nd	measured in months (with fractions permitted)	
numpredoutformal	Number of different criteria (dependent variables) for which predictions of human behavior have been made *through a formal process* prior to study participation	IV	2nd	+1 for each	
numpredouttraining	Number of different criteria (dependent variables) for which predictions of human behavior have been made *after prediction training was received* (prior to study participation)	IV	2nd	+1 for each	
Numpredouttrainingformal	Number of different criteria (dependent variables) for which predictions of human behavior have been made after *formal* prediction training was received (prior to study participation)	IV	2nd	+1 for each	
lengthtraining	Aggregate length of time over which any training to make predictions (judgments) was received	IV	2nd	measured in weeks (with fractions permitted)	
amtstats	Number of statistics courses taken	IV	2nd	+.5 = HS class; +1 = college class; +2 = grad school class)	
amtdecisionmkg	Number of judgment and decision making courses taken	IV	2nd	+.5 = HS class; +1 = college class; +2 = grad school class)	

[84] All 1st level (within-subjects variables) for a multilevel analysis can act as 2nd-level (between-subjects variables) for another context. For example, they act as 2nd-level variables in the intercorrelation matrix in Table 10 and Table 19.

Table 5 – *cont'd.*

Variable Name	Variable Description	Dependent or Independent Variable (DV or IV)	Level (1st = Within Subjects; 2nd = Between Subjects)[85]	Additional Scale Information (or N/A If Variable Was Not Analyzed)	
				Fall 2009 Data (4 blocks; 100 predictions)	Spring 2010 Data (5 blocks; 125 predictions)
• *Insight:*					
PerExprAware	Percentage of blocks during which awareness of the disordinal interaction was expressed	IV	2nd	Proportion	
PerSaidUsed	Percentage of blocks during which the subject explicitly replied "Yes" to a question asking whether the subject used the disordinal interaction	IV	2nd	Proportion	
PerUsedCorrectly	Percentage of blocks during which the subject correcty used the disordinal interaction	IV	2nd	Proportion	
PerSaidUsed3rdVar	Percentage of blocks during which the subject indicated that s/he used a variable not described in the study	IV	2nd	Proportion	
TrPerExprAware	Arcsine transformed PerExprAware (to normalize the variable)	IV	2nd	2 * arcsine(Proportion^.5)	
TrPerSaidUsed	Arcsine transformed PerSaidUsed (to normalize the variable)	IV	2nd	2 * arcsine(Proportion^.5)	
TrPerUsedCorrectly	Arcsine transformed PerUsedCorrectly (to normalize the variable)	IV	2nd	2 * arcsine(Proportion^.5)	
TrPerSaidUsed3rdVar	Arcsine transformed PerSaidUsed3rdVar (to normalize the variable)	IV	2nd	2 * arcsine(Proportion^.5)	

[85] All 1st level (within-subjects variables) for a multilevel analysis can act as 2nd-level (between-subjects variables) for another context. For example, they act as 2nd-level variables in the intercorrelation matrix in Table 10 and Table 19.

Table 6. *Mechanical Versus Clinical Criterion-Related Accuracy Analyses (Fall 2009)*

	Mechanical		Clinical					
	Accuracy	Zp of Accuracy	Feedback Group	Mean Estimate	95% CI of Mean Estimate	Zr of Mean Estimate	N of Mean Estimate	z (Sig. if > 1.96)
Criterion-Related Validity (R_e or r_a)	0.168	--	Group 1	0.118	--	--	--	--
	0.168	0.18	Group 2	0.173	[-0.134, 0.449]	0.175	43	0.316
	0.168	--	Group 3	0.154	--	--	--	--
Skill Score	0.0102	--	Group 1	-1.841	--	--	--	--
	0.0102	--	Group 2	-1.464	--	--	--	--
	0.0102	--	Group 3	-1.29	--	--	--	--

Table 7. *Changes Over Time in Accuracy and the Determinants of Accuracy (Fall 2009)*

	Fisher r_a			*Skill Score*			*Fisher C*			*Fisher r_z*		
	Estimate	p-value or 95% Confidence Interval	N_{obs} (N_{sub})	Estimate	p-value or 95% Confidence Interval	N_{obs} (N_{sub})	Estimate	p-value or 95% Confidence Interval	N_{obs} (N_{sub})	Estimate	p-value or 95% Confidence Interval	N_{obs} (N_{sub})
Step 1: Fit of Random Intercepts												
ICC:	0.27	--	568 (142)	0.35	--	568 (142)	0.23	--	568 (142)	0.23	--	568 (142)
Likelihood Ratio χ^2 (where df = 1):	50.086	< 0.0001	568 (142)	80.814	< 0.0001	568 (142)	37.727	< 0.0001	568 (142)	37.672	< 0.0001	568 (142)
Step 2: Fit of Random Slopes												
Log likelihood χ^2 (where df = 2):	21.572	0.00002	568 (142)	7.912	0.0191	568 (142)	5.791	0.055	568 (142)	5.804	0.055	568 (142)
Step 3: Fit of Interaction (Feedback Condition x Time)												
Log likelihood χ^2 or F (where df = 4):	3.941	0.414	568 (142)	11.739	0.0194	568 (142)	9.352	0.053	568 (142)	9.404	0.052	568 (142)
Step 4: Slope(s) (unstandardized *γ* or *b* values) of Growth Curve(s)												
Overall (Fixed Effect):	-0.006	[-0.017, 0.005]	568 (142)	0.012	[-0.074, 0.097]	568 (142)	-0.008	[-0.026, 0.010]	568 (142)	-0.008	[-0.025, 0.010]	568 (142)
Feedback Group 1:	--	--	--	-0.134	[-0.254, -0.014]	192 (48)	-0.041	[-0.068, -0.014]	192 (48)	-0.044	[-0.067, -0.013]	192 (48)
Feedback Group 2:	--	--	--	0.032	[-0.095, 0.158]	172 (43)	0.018	[-0.011, 0.047]	172 (43)	0.018	[-0.011, 0.046]	172 (43)
Feedback Group 3:	--	--	--	0.132	[0.016, 0.248]	204 (51)	0.001	[-0.025, 0.028]	204 (51)	0.001	[-0.025, 0.027]	204 (51)

Notes. N_{obs} = number of observations based upon which the estimate was calculated; N_{sub} = number of subjects based upon which the estimate was calculated; df = degrees of freedom

Table 7 – *cont'd.*

	Fisher G			*Fisher R_s*			*C_{xj} (Cwtx1x2_s)*		
	Estimate	p-value or 95% Confidence Interval	N_{obs} (N_{sub})	Estimate	p-value or 95% Confidence Interval	N_{obs} (N_{sub})	Estimate	p-value or 95% Confidence Interval	N_{obs} (N_{sub})
Step 1: Fit of Random Intercepts									
ICC:	0.26	--	568 (142)	0.27	--	568 (142)	0.35	--	568 (142)
Likelihood Ratio χ^2 (where df = 1):	45.646	< 0.0001	568 (142)	4.574	< 0.0001	568 (142)	82.11	< 0.0001	568 (142)
Step 2: Fit of Random Slopes									
Log likelihood χ^2 (where df = 2):	0.534	0.766	568 (142)	48.924	$2.378*10^{-11}$	568 (142)	23.691	$7.172*10^{-6}$	568 (142)
Step 3: Fit of Interaction (Feedback Condition x Time)									
Log likelihood χ^2 or F (where df = 4):	1.297	0.862	568 (142)	14.687	0.005	568 (142)	0.927	0.921	568 (142)
Step 4: Slope(s) (unstandardized *γ* or *b* values) of Growth Curve(s)									
Overall (Fixed Effect):	-0.048	[-0.135, 0.039]	568 (142)	0.043	[-0.038, 0.123]	568 (142)	-0.0014	[-0.009, 0.007]	568 (142)
Feedback Group 1:	--	--	--	0.207	[0.074, 0.340]	192 (48)	--	--	--
Feedback Group 2:	--	--	--	0.005	[-0.136, 0.146]	172 (43)	--	--	--
Feedback Group 3:	--	--	--	-0.08	[-0.209, 0.049]	204 (51)	--	--	--

Notes. N_{obs} = number of observations based upon which the estimate was calculated; N_{sub} = number of subjects based upon which the estimate was calculated; df = degrees of freedom

Table 8. *Changes Over Time in Overall Confidence (Fall 2009 and Spring 2010)*

	d	**95% CI for *d***	***N* for *d***	**mean difference**	***t***	***df***	***p*-value**
Fall 2009:							
Absolute Confidence							
Group 1	-0.25	[-.72, .23]	30	-0.2333	1.07	29	0.293
Group 2	-0.82	[-1,27, -0.37]	34	-0.7059	3.69	33	0.0008
Group 3	-0.61	[-0.87, -0.35]	38	-0.6974	4.71	37	$3.401*10^{-5}$
Relative Confidence							
Group 1	-0.4	[-0.81, 0.02]	28	-0.3214	1.97	27	0.06
Group 2	-0.48	[-0.89, -0.06]	35	-0.4286	2.32	34	0.03
Group 3	-0.48	[-0.75, -0.22]	37	-0.5676	3.72	36	0.0007
Spring 2010:							
Absolute Confidence							
Group 1	-0.76	[-1.26, -0.26]	33	-0.606	3.12	32	0.004
Group 2	-0.58	[-0.965, -0.196]	41	-0.512	3.05	40	0.004
Group 3	-0.6	[-0.99, -0.21]	40	-0.4625	3.1	39	0.004
Relative Confidence							
Group 1	-0.43	[-0.92, 0.07]	32	-0.375	1.75	31	0.09
Group 2	-0.42	[-0.71, -0.13]	41	-0.415	2.97	40	0.005
Group 3	-0.5	[-0.75, -0.25]	41	-0.512	4.05	40	0.0002

Notes. *d* = Cohen's *d*, where a negative value indicates a decline in overall confidence between the start of the study and the end of the study; 95% CI for *d* = 95% confidence interval for the Cohen's *d*; *N* for *d* = sample size for the Cohen's *d*; mean difference = mean of the differences for that group

Table 9. *Changes Over Time in Confidence Measured After Every Block (Fall 2009 and Spring 2010)*

	Fall 2009 Absolute Confidence			Fall 2009 Relative Confidence			Spring 2010 Absolute Confidence			Spring 2010 Relative Confidence		
	Estimate	p-value or 95% Confidence Interval	N_{obs} (N_{sub})	Estimate	p-value or 95% Confidence Interval	N_{obs} (N_{sub})	Estimate	p-value or 95% Confidence Interval	N_{obs} (N_{sub})	Estimate	p-value or 95% Confidence Interval	N_{obs} (N_{sub})
Step 1: Fit of Random Intercepts												
ICC:	0.64	--	533 (137)	0.71	--	527 (136)	0.57	--	691 (141)	0.66	--	691 (141)
Likelihood Ratio (where df = 1):	270.772	< 0.0001	533 (137)	339.884	< 0.0001	527 (136)	284.162	< 0.0001	691 (141)	414.05	< 0.0001	691 (141)
Step 2: Fit of Random Slopes												
Log likelihood χ^2 (where df = 2):	23.308	$8.685*10^{-6}$	533 (137)	33.51	$5.29*10^{-8}$	527 (136)	26.596	$1.678*10^{-6}$	691 (141)	19.808	$4.998*10^{-5}$	691 (141)
Step 3: Fit of Interaction (Feedback Condition x Time)												
Log likelihood χ^2 or F (where df = 4):	11.89	0.018	533 (137)	8.372	0.079	527 (136)	8.724	0.068	691 (141)	7.2	0.126	691 (141)
Step 4: Slope(s) (unstandardized *γ* or *b* values) of Growth Curve(s)												
Overall (Fixed Effect):	-0.047	[-0.131, 0.037]	533 (137)	-0.052	[-0.108, -0.009]	527 (136)	0.017	[-0.025, 0.060]	691 (141)	-0.01	[-0.044, 0.026]	691 (141)
Feedback Group 1:	0.09	[-0.010, 0.190]	183 (48)	0.05	[-0.054, 0.112]	183 (48)	0.018	[-0.054, 0.096]	222 (46)	--	--	--
Feedback Group 2:	-0.154	[-0.260, -0.048]	158 (40)	-0.11	[-0.201, -0.025]	158 (40)	-0.03	[-0.101, 0.044]	237 (48)	--	--	--
Feedback Group 3:	-0.074	[-0.171, 0.024]	192 (49)	-0.104	[-0.179, -0.014]	186 (48)	0.063	[-0.011, 0.134]	232 (47)	--	--	--

Notes. N_{obs} = number of observations based upon which the estimate was calculated; N_{sub} = number of subjects based upon which the estimate was calculated; df = degrees of freedom

Table 10. *Correlations Among Second-Level (Between-Persons) Variables (Fall 2009)*

	Fisherra	Fisherra1	Fisherra2	Fisherra3	Fisherra4	SS	SS1	SS2	SS3	SS4	FisherRslin	FisherRslin1	FisherRslin2	FisherRslin3	FisherRslin4
Fisherra	1**	0.54**	0.61**	0.73**	0.73**	0.47**	0.31**	0.3**	0.4**	0.34**	-0.13	-0.26**	-0.12	-0.13	-0.09
Fisherra1	0.54**	1**	0.25**	0.21*	0.07	0.24**	0.54**	0.04	0.13	0	-0.11	-0.47**	-0.02	-0.05	-0.06
Fisherra2	0.61**	0.25**	1**	0.32**	0.19*	0.36**	0.25**	0.35**	0.32**	0.18*	-0.09	-0.25**	-0.2*	-0.17*	-0.06
Fisherra3	0.73**	0.21*	0.32**	1**	0.5**	0.32**	0.06	0.19*	0.44**	0.25**	-0.1	-0.16	-0.01	-0.1	-0.07
Fisherra4	0.73**	0.07	0.19*	0.5**	1**	0.36**	0.01	0.25**	0.27**	0.46**	-0.06	0.05	-0.07	-0.03	-0.09
SS	0.47**	0.24**	0.36**	0.32**	0.36**	1**	0.53**	0.84**	0.8**	0.79**	-0.13	-0.04	-0.08	-0.13	-0.13
SS1	0.31**	0.54**	0.25**	0.06	0.01	0.53**	1**	0.24**	0.32**	0.06	-0.01	-0.22**	0	0.07	0.12
SS2	0.3**	0.04	0.35**	0.19*	0.25**	0.84**	0.24**	1**	0.55**	0.7**	-0.13	0.04	-0.16	-0.24**	-0.12
SS3	0.4**	0.13	0.32**	0.44**	0.27**	0.8**	0.32**	0.55**	1**	0.53**	-0.12	-0.05	-0.01	-0.13	-0.09
SS4	0.34**	0	0.18*	0.25**	0.46**	0.79**	0.06	0.7**	0.53**	1**	-0.08	0.13	-0.03	-0.1	-0.28**
FisherRslin	-0.13	-0.11	-0.09	-0.1	-0.06	-0.13	-0.01	-0.13	-0.12	-0.08	1**	0.55**	0.43**	0.44**	0.31**
FisherRslin1	-0.26**	-0.47**	-0.25**	-0.16	0.05	-0.04	-0.22**	0.04	-0.05	0.13	0.55**	1**	0.25**	0.16	0.11
FisherRslin2	-0.12	-0.02	-0.2*	-0.01	-0.07	-0.08	0	-0.16	-0.01	-0.03	0.43**	0.25**	1**	0.29**	0.27**
FisherRslin3	-0.13	-0.05	-0.17*	-0.1	-0.03	-0.13	0.07	-0.24**	-0.13	-0.1	0.44**	0.16	0.29**	1**	0.5**
FisherRslin4	-0.09	-0.06	-0.06	-0.07	-0.09	-0.13	0.12	-0.12	-0.09	-0.28**	0.31**	0.11	0.27**	0.5**	1**
FisherG	0.14	0.1	0.06	0.05	0.1	0.14	0.24**	0.05	0.06	0.09	0.35**	0.27**	0.22**	0.16	0.19*
FisherG1	-0.04	0.04	0	-0.1	-0.06	0.17*	0.33**	0.05	0.07	0.09	0.44**	0.41**	0.23**	0.28**	0.18*
FisherG2	0.27**	0.18*	0.35**	0.13	0.13	0.25**	0.19*	0.17*	0.18*	0.21*	0.13	0.04	-0.04	-0.09	-0.03
FisherG3	0.21*	0.06	0.11	0.29**	0.14	0.29**	0.19*	0.2*	0.26**	0.23**	0.27**	0.13	0.27**	0.15	-0.03
FisherG4	0.25**	0.06	0.03	0.18*	0.3**	0.24**	0.21*	0.1	0.21*	0.18*	0.45**	0.18*	0.21*	0.2*	0.17*

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 10 – *cont'd.*

	Fisherra	Fisherra1	Fisherra2	Fisherra3	Fisherra4	SS	SS1	SS2	SS3	SS4	FisherRslin	FisherRslin1	FisherRslin2	FisherRslin3	FisherRslin4
FisherCstat	0.84**	0.47**	0.52**	0.59**	0.61**	0.26**	0.14	0.17*	0.2*	0.19*	-0.26**	-0.34**	-0.21*	-0.19*	-0.16
FisherCstat1	0.41**	0.78**	0.16	0.17*	0.06	0.05	0.26**	-0.05	0	-0.07	-0.19*	-0.48**	-0.04	-0.08	-0.08
FisherCstat2	0.5**	0.2*	0.84**	0.26**	0.14	0.25**	0.15	0.29**	0.23**	0.11	-0.1	-0.22**	-0.14	-0.1	0.01
FisherCstat3	0.58**	0.17*	0.3**	0.83**	0.35**	0.16	0.02	0.04	0.3**	0.09	-0.24**	-0.22**	-0.04	-0.19*	-0.12
FisherCstat4	0.62**	0.1	0.14	0.39**	0.86**	0.29**	-0.06	0.24**	0.16	0.44**	-0.21*	-0.02	-0.18*	-0.16	-0.26**
Fisherrz	0.84**	0.47**	0.52**	0.59**	0.61**	0.26**	0.14	0.17*	0.2*	0.19*	-0.26**	-0.34**	-0.21*	-0.19*	-0.16
Fisherrz1	0.41**	0.78**	0.16	0.17*	0.06	0.05	0.26**	-0.05	0	-0.07	-0.19*	-0.48**	-0.04	-0.08	-0.08
Fisherrz2	0.5**	0.19*	0.84**	0.25**	0.14	0.25**	0.15	0.29**	0.23**	0.11	-0.1	-0.22**	-0.14	-0.11	0.01
Fisherrz3	0.58**	0.17*	0.3**	0.83**	0.35**	0.16	0.02	0.04	0.3**	0.09	-0.24**	-0.22**	-0.04	-0.19*	-0.12
Fisherrz4	0.62**	0.1	0.14	0.39**	0.86**	0.29**	-0.06	0.24**	0.16	0.44**	-0.21*	-0.02	-0.18*	-0.16	-0.26**
Cwtx1x2_s	0.31**	0.17*	0.21*	0.2*	0.24**	0.2*	0.23**	0.09	0.15	0.14	0.79**	0.36**	0.35**	0.3**	0.19*
Cwtx1x2_1_s	0.31**	0.13	0.06	0.25**	0.32**	0.13	-0.04	0.1	0.17*	0.14	0	0.12	0.06	-0.1	-0.1
Cwtx1x2_2_s	0.19*	0.15	0.27**	0.09	0.07	0.06	0.11	-0.1	0.15	0.03	0.39**	0.13	0.4**	0.16	0.04
Cwtx1x2_3_s	0.17*	-0.04	-0.06	0.31**	0.3**	0.01	-0.08	-0.07	0.08	0.05	0.37**	0.18*	0.21*	0.4**	0.13
Cwtx1x2_4_s	0.32**	-0.03	0.14	0.25**	0.42**	0.12	0.01	0.03	0.2*	0.08	0.12	0.14	0	0.1	0.13
ConfAbsOverallPre	-0.15	-0.23*	0	-0.14	-0.04	-0.14	-0.14	-0.02	-0.17	-0.07	0.15	0.24*	0.04	0.03	0.03
ConfAbs1	-0.08	-0.2*	0.03	-0.03	-0.02	-0.17*	-0.19*	-0.03	-0.17*	-0.08	0.09	0.17*	0.02	0.02	-0.05
ConfAbs2	0.01	-0.01	0.1	-0.09	0.03	-0.22*	-0.09	-0.18*	-0.19*	-0.15	0.21*	0.04	0.16	0.03	-0.01
ConfAbs3	-0.06	0.05	0	-0.11	-0.07	-0.22*	-0.01	-0.19*	-0.17	-0.23**	0.23**	0	0.07	0.07	0.06
ConfAbs4	-0.1	-0.01	0.05	-0.21*	-0.12	-0.27**	-0.03	-0.18*	-0.2*	-0.31**	0.13	-0.03	0.08	0.01	0.1
ConfAbsOverallPost	0	0.03	0.07	-0.08	-0.05	-0.23**	0	-0.19*	-0.19*	-0.22*	0.2*	-0.02	0.11	0.05	0.01

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 10 – *cont'd.*

	Fisherra	Fisherra1	Fisherra2	Fisherra3	Fisherra4	SS	SS1	SS2	SS3	SS4	FisherRslin	FisherRslin1	FisherRslin2	FisherRslin3	FisherRslin4
ConfRelOverallPre	-0.06	-0.1	0.11	-0.16	-0.04	-0.03	-0.03	0.09	-0.11	-0.01	0.08	0.04	0.04	-0.02	-0.02
ConfRel1	-0.11	-0.14	0.02	-0.07	-0.08	-0.21*	-0.21*	-0.04	-0.19*	-0.14	0.16	0.12	0.07	-0.01	0.03
ConfRel2	-0.13	-0.12	-0.06	-0.15	-0.02	-0.24**	-0.18*	-0.16	-0.22**	-0.12	0.26**	0.11	0.16	0.04	0.07
ConfRel3	-0.13	-0.02	-0.02	-0.15	-0.16	-0.19*	-0.09	-0.12	-0.16	-0.15	0.2*	0.02	0.06	0.04	-0.01
ConfRel4	-0.13	-0.02	0.04	-0.25**	-0.16	-0.24**	-0.05	-0.14	-0.2*	-0.24**	0.17	0.03	0.13	0.02	0.12
ConfRelOverallPost	-0.06	-0.04	0.02	-0.17*	-0.03	-0.13	-0.02	-0.06	-0.15	-0.13	0.14	0	0	0.09	0.06
ACT_COMP_SCR	0.32**	-0.07	0.27**	0.28**	0.36**	0.25**	-0.05	0.27**	0.22*	0.3**	-0.1	0.11	-0.08	-0.13	-0.09
ACT_ENGL_SCR	0.38**	0.07	0.33**	0.27**	0.34**	0.23*	0.01	0.23*	0.22*	0.23*	-0.09	0.11	-0.18	-0.15	-0.08
ACT_ENGWR_SCR	0.39**	0.09	0.34**	0.29**	0.3**	0.26**	-0.01	0.25*	0.26**	0.26**	-0.14	0.06	-0.23*	-0.23*	-0.13
ACT_MATH_SCR	0.2*	-0.17	0.18	0.24*	0.28**	0.22*	-0.08	0.23*	0.23*	0.25**	-0.04	0.13	-0.15	0	0.02
ACT_READ_SCR	0.28**	-0.02	0.22*	0.21*	0.32**	0.22*	-0.05	0.25**	0.14	0.3**	-0.14	0.05	-0.04	-0.18	-0.17
ACT_SCIRE_SCR	0.24*	-0.18	0.27**	0.21*	0.34**	0.22*	-0.04	0.26**	0.12	0.26**	-0.06	0.08	-0.17	-0.13	-0.08
CUM_GPA	0.24**	0.04	0.11	0.18*	0.23**	0.11	-0.05	0.11	0.1	0.18*	0.07	0.09	0	-0.07	-0.09
TOT_ACAD_HOURS	0.09	0.11	0.15	0.05	-0.05	0.12	0.04	0.12	0.13	0.06	0.05	0.01	-0.06	-0.12	-0.07
HS_RANK_PCT	0.2*	0.17	0.3**	-0.06	0.11	0.1	0.08	0.05	0.05	0.08	0.03	0.04	0.02	-0.06	0.01
Gender	0.02	-0.16	0	0.07	0.1	0.05	-0.01	0.08	-0.02	0.08	-0.02	0.02	-0.14	-0.1	-0.08

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 10 – *cont'd.*

	Fisherra	Fisherra1	Fisherra2	Fisherra3	Fisherra4	SS	SS1	SS2	SS3	SS4	FisherRslin	FisherRslin1	FisherRslin2	FisherRslin3	FisherRslin4
extroGoldberg	-0.16	-0.06	-0.23**	-0.19*	-0.05	-0.12	-0.04	-0.11	-0.08	-0.1	0.07	0.12	0.16*	-0.06	0
neurotGoldberg	-0.13	-0.05	-0.03	-0.13	-0.09	-0.06	0.03	-0.05	-0.06	-0.11	0.02	-0.08	-0.07	0.12	0.06
intelGoldberg	0.07	0.1	0.18*	-0.08	0.01	-0.03	-0.01	-0.1	-0.01	0.03	-0.04	-0.08	0.1	0.06	0.07
agreeGoldberg	0.04	0.14	-0.1	0.05	0.03	-0.11	-0.07	-0.16	-0.04	-0.05	0.13	-0.02	0.14	0.08	0.1
conGoldberg	-0.06	0.04	-0.06	-0.03	-0.06	-0.08	-0.04	-0.06	-0.07	-0.04	0.15	0.02	0.11	0.13	0
pvIPC7	-0.14	-0.07	-0.02	-0.13	-0.14	-0.18*	0	-0.11	-0.24**	-0.17*	0.05	0.01	-0.07	0.02	0.01
nvIPC7	-0.04	0.04	0.01	-0.09	-0.08	0.01	0.1	0.01	-0.04	-0.03	-0.15	-0.07	-0.06	-0.14	-0.06
pemIPC7	-0.13	0	-0.24**	-0.15	-0.04	-0.14	-0.05	-0.15	-0.06	-0.13	0.04	0.09	0.12	-0.14	0
nemIPC7	-0.07	-0.04	-0.04	-0.03	-0.03	-0.05	-0.05	-0.07	-0.04	-0.03	0	-0.05	-0.02	0.08	0.04
conIPC7	-0.17*	0.11	-0.14	-0.15	-0.2*	-0.19*	-0.07	-0.07	-0.22**	-0.14	0.04	-0.02	-0.09	0.06	-0.05
agreeIPC7	0.13	0.23**	-0.01	0.07	0.07	0.04	0.04	-0.01	0.04	0.04	-0.04	0.02	0.09	-0.01	0.01
cnvIPC7	-0.14	-0.13	-0.17*	-0.03	-0.05	-0.22**	-0.14	-0.12	-0.24**	-0.15	0.14	0.12	-0.06	0.15	0.03
ReaRIASEC	-0.06	-0.14	0.03	0.02	-0.04	0.18*	0.08	0.17*	0.17*	0.1	-0.02	0.09	-0.04	0.04	0.14
InvRIASEC	0.03	-0.12	0.18*	0.08	0.02	0.17*	0.02	0.16	0.16	0.16	0.07	0.17*	-0.1	0.07	0.14
ArtRIASEC	0.01	0	0.11	-0.01	-0.03	0.09	0.01	0.02	0.18*	0.05	-0.12	0.05	0.1	0.03	0.09
SocRIASEC	-0.11	-0.02	-0.13	-0.05	-0.11	0.07	0.11	0.02	0.07	0.02	0.15	0.12	0.05	0.12	0.11
EntRIASEC	-0.18*	0.07	-0.23**	-0.16	-0.15	-0.08	0.14	-0.08	-0.07	-0.18*	-0.03	-0.08	0.04	0.06	0.04
ConRIASEC	0.2*	-0.06	-0.14	-0.15	-0.14	-0.05	0.09	0	-0.11	-0.13	0.14	0.14	0.04	0.04	0

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 10 – *cont'd.*

	Fisherra	Fisherra1	Fisherra2	Fisherra3	Fisherra4	SS	SS1	SS2	SS3	SS4	FisherRslin	FisherRslin1	FisherRslin2	FisherRslin3	FisherRslin4
numpredout	-0.01	-0.06	0.08	0.05	-0.01	0.1	0.13	0.05	0.11	0.02	-0.01	-0.09	-0.07	0.03	-0.01
numratees	0.12	0.06	0.09	0.14	0.09	0.15	0.19*	0.05	0.12	0.08	-0.06	-0.04	-0.01	0	-0.09
lengthtimepred	-0.02	-0.11	0.01	0.09	-0.02	-0.07	0.02	-0.1	0.01	-0.11	-0.03	-0.11	-0.04	-0.03	-0.05
numpredoutformal	0.03	0.11	0.08	0.04	-0.09	0.14	0.21*	0.12	0.08	0	-0.14	-0.14	-0.13	-0.11	-0.1
numpredouttraining	-0.06	0.04	0.07	-0.01	-0.11	0.01	0.04	0.01	0.08	-0.06	-0.03	-0.14	-0.03	0.02	-0.03
numpredouttrainingformal	-0.11	-0.09	-0.07	-0.04	-0.08	-0.14	-0.01	-0.16	-0.06	-0.17*	0.05	-0.07	-0.02	0.04	0.05
lengthtraining	-0.16	-0.07	-0.19*	-0.1	-0.11	-0.06	0.04	-0.06	-0.09	-0.09	-0.12	-0.14	-0.13	-0.07	-0.07
amtstats	0	0.05	0.04	-0.09	-0.02	0.02	0.07	0.06	-0.02	-0.06	0.04	-0.02	-0.09	-0.05	0
amtdecisionmkg	-0.15	0.06	-0.2*	-0.16	-0.13	-0.03	0.14	-0.05	-0.05	-0.15	-0.08	-0.07	-0.1	-0.07	-0.05
TrPerExprAware	0.13	0.16	0.09	0	0.08	0	0.01	-0.03	0	0.02	-0.15	-0.21*	-0.17*	-0.1	-0.12
TrPerSaidUsed	0.15	0.11	0.09	0.06	0.13	0.05	-0.05	0.02	0.05	0.11	-0.22**	-0.2*	-0.18*	-0.14	-0.19*
TrPerUsedCorrectly	0.18*	0.08	0.14	0.13	0.12	0.06	0.04	-0.04	0.1	0.1	-0.04	-0.04	-0.04	-0.1	-0.13
TrPerSaidUsed3rdVar	0.04	0.12	0.06	0.06	-0.04	0.08	0.06	0.06	0.07	0.06	0.01	-0.07	-0.03	-0.03	-0.02

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 10 – *cont'd.*

	FisherG	FisherG1	FisherG2	FisherG3	FisherG4	FisherCstat	FisherCstat1	FisherCstat2	FisherCstat3	FisherCstat4	Fisherrz	Fisherrz1	Fisherrz2	Fisherrz3	Fisherrz4
Fisherra	0.14	-0.04	0.27**	0.21*	0.25**	0.84**	0.41**	0.5**	0.58**	0.62**	0.84**	0.41**	0.5**	0.58**	0.62**
Fisherra1	0.1	0.04	0.18*	0.06	0.06	0.47**	0.78**	0.2*	0.17*	0.1	0.47**	0.78**	0.19*	0.17*	0.1
Fisherra2	0.06	0	0.35**	0.11	0.03	0.52**	0.16	0.84**	0.3**	0.14	0.52**	0.16	0.84**	0.3**	0.14
Fisherra3	0.05	-0.1	0.13	0.29**	0.18*	0.59**	0.17*	0.26**	0.83**	0.39**	0.59**	0.17*	0.25**	0.83**	0.39**
Fisherra4	0.1	-0.06	0.13	0.14	0.3**	0.61**	0.06	0.14	0.35**	0.86**	0.61**	0.06	0.14	0.35**	0.86**
SS	0.14	0.17*	0.25**	0.29**	0.24**	0.26**	0.05	0.25**	0.16	0.29**	0.26**	0.05	0.25**	0.16	0.29**
SS1	0.24**	0.33**	0.19*	0.19*	0.21*	0.14	0.26**	0.15	0.02	-0.06	0.14	0.26**	0.15	0.02	-0.06
SS2	0.05	0.05	0.17*	0.2*	0.1	0.17*	-0.05	0.29**	0.04	0.24**	0.17*	-0.05	0.29**	0.04	0.24**
SS3	0.06	0.07	0.18*	0.26**	0.21*	0.2*	0	0.23**	0.3**	0.16	0.2*	0	0.23**	0.3**	0.16
SS4	0.09	0.09	0.21*	0.23**	0.18*	0.19*	-0.07	0.11	0.09	0.44**	0.19*	-0.07	0.11	0.09	0.44**
FisherRslin	0.35**	0.44**	0.13	0.27**	0.45**	-0.26**	-0.19*	-0.1	-0.24**	-0.21*	-0.26**	-0.19*	-0.1	-0.24**	-0.21*
FisherRslin1	0.27**	0.41**	0.04	0.13	0.18*	-0.34**	-0.48**	-0.22**	-0.22**	-0.02	-0.34**	-0.48**	-0.22**	-0.22**	-0.02
FisherRslin2	0.22**	0.23**	-0.04	0.27**	0.21*	-0.21*	-0.04	-0.14	-0.04	-0.18*	-0.21*	-0.04	-0.14	-0.04	-0.18*
FisherRslin3	0.16	0.28**	-0.09	0.15	0.2*	-0.19*	-0.08	-0.1	-0.19*	-0.16	-0.19*	-0.08	-0.11	-0.19*	-0.16
FisherRslin4	0.19*	0.18*	-0.03	-0.03	0.17*	-0.16	-0.08	0.01	-0.12	-0.26**	-0.16	-0.08	0.01	-0.12	-0.26**
FisherG	1**	0.63**	0.32**	0.38**	0.57**	-0.2*	-0.2*	-0.05	-0.09	-0.09	-0.2*	-0.2*	-0.04	-0.09	-0.09
FisherG1	0.63**	1**	0.3**	0.32**	0.34**	-0.33**	-0.31**	-0.08	-0.18*	-0.21*	-0.33**	-0.31**	-0.08	-0.18*	-0.21*
FisherG2	0.32**	0.3**	1**	0.06	0.25**	0.03	-0.03	0.02	0.04	0	0.03	-0.03	0.02	0.04	0
FisherG3	0.38**	0.32**	0.06	1**	0.27**	-0.08	-0.14	0.06	0	0	-0.08	-0.14	0.06	0	0
FisherG4	0.57**	0.34**	0.25**	0.27**	1**	-0.04	-0.14	-0.07	0.03	0.05	-0.04	-0.14	-0.07	0.03	0.05

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 10 – *cont'd.*

	FisherG	FisherG1	FisherG2	FisherG3	FisherG4	FisherCstat	FisherCstat1	FisherCstat2	FisherCstat3	FisherCstat4	Fisherrz	Fisherrz1	Fisherrz2	Fisherrz3	Fisherrz4
FisherCstat	-0.2*	-0.33**	0.03	-0.08	-0.04	1**	0.58**	0.54**	0.62**	0.71**	1**	0.58**	0.54**	0.62**	0.71**
FisherCstat1	-0.2*	-0.31**	-0.03	-0.14	-0.14	0.58**	1**	0.17*	0.19*	0.2*	0.58**	1**	0.17*	0.19*	0.2*
FisherCstat2	-0.05	-0.08	0.02	0.06	-0.07	0.54**	0.17*	1**	0.27**	0.16	0.54**	0.17*	1**	0.27**	0.16
FisherCstat3	-0.09	-0.18*	0.04	0	0.03	0.62**	0.19*	0.27**	1**	0.36**	0.62**	0.19*	0.27**	1**	0.36**
FisherCstat4	-0.09	-0.21*	0	0	0.05	0.71**	0.2*	0.16	0.36**	1**	0.71**	0.2*	0.16	0.36**	1**
Fisherrz	-0.2*	-0.33**	0.03	-0.08	-0.04	1**	0.58**	0.54**	0.62**	0.71**	1**	0.58**	0.54**	0.62**	0.71**
Fisherrz1	-0.2*	-0.31**	-0.03	-0.14	-0.14	0.58**	1**	0.17*	0.19*	0.2*	0.58**	1**	0.17*	0.19*	0.2*
Fisherrz2	-0.04	-0.08	0.02	0.06	-0.07	0.54**	0.17*	1**	0.27**	0.16	0.54**	0.17*	1**	0.27**	0.16*
Fisherrz3	-0.09	-0.18*	0.04	0	0.03	0.62**	0.19*	0.27**	1**	0.36**	0.62**	0.19*	0.27**	1**	0.36**
Fisherrz4	-0.09	-0.21*	0	0	0.05	0.71**	0.2*	0.16	0.36**	1**	0.71**	0.2*	0.16*	0.36**	1**
Cwtx1x2_s	0.54**	0.48**	0.35**	0.41**	0.63**	0.03	-0.07	0.09	-0.01	0.01	0.03	-0.07	0.09	-0.01	0.01
Cwtx1x2_1_s	-0.07	-0.13	-0.04	-0.05	-0.05	0.32**	0.15	0.09	0.19*	0.29**	0.32**	0.15	0.09	0.19*	0.29**
Cwtx1x2_2_s	0.08	0.18*	0.11	0.18*	0.14	0.12	0.07	0.21*	0.09	-0.02	0.12	0.07	0.21*	0.09	-0.02
Cwtx1x2_3_s	0.1	0.08	-0.03	0.15	0.24**	0.13	0.01	-0.02	0.18*	0.15	0.13	0.01	-0.02	0.18*	0.15
Cwtx1x2_4_s	-0.04	0.01	0.12	-0.01	0.15	0.28**	-0.02	0.14	0.18*	0.27**	0.28**	-0.02	0.13	0.18*	0.27**
ConfAbsOverallPre	-0.01	-0.04	-0.15	-0.11	-0.01	-0.12	-0.15	0.02	-0.14	-0.03	-0.12	-0.15	0.02	-0.14	-0.03
ConfAbs1	-0.1	-0.1	-0.18*	-0.01	-0.15	-0.03	-0.19*	0.08	-0.01	0	-0.03	-0.19*	0.08	-0.01	0
ConfAbs2	-0.04	-0.06	0	0.06	-0.02	-0.01	0.01	0.04	-0.05	-0.02	-0.01	0.01	0.04	-0.05	-0.02
ConfAbs3	0	-0.05	0.03	-0.01	0.02	-0.1	0.01	-0.01	-0.08	-0.15	-0.1	0.01	-0.01	-0.08	-0.15
ConfAbs4	-0.04	-0.07	-0.02	-0.05	-0.12	-0.09	0.01	0.06	-0.18*	-0.15	-0.09	0.01	0.06	-0.18*	-0.15
ConfAbsOverallPost	0.04	-0.11	-0.02	0.07	-0.03	-0.08	-0.04	0.03	-0.1	-0.1	-0.08	-0.04	0.03	-0.1	-0.1

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 10 – *cont'd.*

	FisherG	FisherG1	FisherG2	FisherG3	FisherG4	FisherCstat	FisherCstat1	FisherCstat2	FisherCstat3	FisherCstat4	Fisherrz	Fisherrz1	Fisherrz2	Fisherrz3	Fisherrz4
ConfRelOverallPre	-0.03	-0.03	-0.11	-0.12	0.05	-0.05	-0.08	0.11	-0.16	-0.04	-0.05	-0.08	0.11	-0.16	-0.04
ConfRel1	-0.03	-0.06	-0.07	-0.05	-0.1	-0.08	-0.13	0.04	-0.09	-0.08	-0.08	-0.13	0.04	-0.09	-0.08
ConfRel2	0.02	-0.05	0	-0.05	0.06	-0.12	-0.12	-0.06	-0.09	-0.05	-0.12	-0.12	-0.06	-0.09	-0.05
ConfRel3	-0.03	-0.13	0.01	-0.06	-0.04	-0.11	0.01	-0.01	-0.12	-0.17	-0.11	0.01	-0.01	-0.12	-0.17
ConfRel4	0.03	-0.07	0.02	-0.08	-0.06	-0.11	-0.02	0.04	-0.21*	-0.18*	-0.11	-0.02	0.04	-0.21*	-0.18*
ConfRelOverallPost	0.01	-0.12	-0.04	-0.04	0.04	-0.1	-0.08	0.01	-0.18*	-0.08	-0.1	-0.08	0.01	-0.18*	-0.08
ACT_COMP_SCR	0.07	-0.15	0.24*	-0.07	0.12	0.26**	-0.03	0.2*	0.16	0.27**	0.26**	-0.03	0.2*	0.16	0.27**
ACT_ENGL_SCR	0.09	-0.11	0.19*	-0.05	0.05	0.31**	0.1	0.27**	0.15	0.28**	0.31**	0.1	0.27**	0.15	0.28**
ACT_ENGWR_SCR	0.12	-0.09	0.24*	-0.09	0.01	0.33**	0.13	0.26**	0.18	0.25*	0.33**	0.13	0.26**	0.18	0.25*
ACT_MATH_SCR	0.07	-0.12	0.17	-0.09	0.15	0.13	-0.16	0.14	0.09	0.2*	0.13	-0.16	0.14	0.09	0.2*
ACT_READ_SCR	0.08	-0.2*	0.22*	-0.06	0.12	0.23*	0.01	0.12	0.16	0.22*	0.23*	0.01	0.12	0.16	0.22*
ACT_SCIRE_SCR	-0.02	-0.14	0.25**	-0.16	0.1	0.22*	-0.14	0.2*	0.11	0.24*	0.22*	-0.14	0.2*	0.11	0.24*
CUM_GPA	0.19*	-0.02	0.12	0.04	0.25**	0.13	0.04	0.08	0.09	0.14	0.13	0.04	0.08	0.09	0.14
TOT_ACAD_HOURS	-0.01	0	0.03	-0.02	-0.06	0.11	0.15	0.13	0.05	-0.05	0.11	0.15	0.13	0.05	-0.05
HS_RANK_PCT	0.09	0.12	0.09	-0.05	0.13	0.17	0.12	0.25*	-0.09	0.08	0.17	0.12	0.25*	-0.09	0.08
Gender	-0.07	0	0.01	-0.09	-0.03	0.07	-0.12	0	0.1	0.13	0.07	-0.12	0	0.1	0.13

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 10 – *cont'd.*

	FisherG	FisherG1	FisherG2	FisherG3	FisherG4	FisherCstat	FisherCstat1	FisherCstat2	FisherCstat3	FisherCstat4	Fisherrz	Fisherrz1	Fisherrz2	Fisherrz3	Fisherrz4
extroGoldberg	-0.14	-0.02	-0.06	-0.01	-0.06	-0.15	-0.03	-0.25**	-0.15	-0.08	-0.15	-0.03	-0.25**	-0.16	-0.08
neurotGoldberg	0.05	0.02	0.06	-0.02	0.03	-0.08	-0.08	-0.05	-0.09	0	-0.08	-0.08	-0.05	-0.09	0
intelGoldberg	-0.05	-0.08	-0.04	-0.09	-0.06	0.11	0.12	0.2*	-0.06	-0.01	0.11	0.12	0.2*	-0.06	-0.01
agreeGoldberg	-0.05	0.08	-0.14	-0.03	0.06	0.04	0.13	-0.05	0.04	-0.01	0.04	0.13	-0.05	0.04	-0.01
conGoldberg	0.06	-0.02	-0.09	-0.02	0.05	-0.06	0.06	-0.05	-0.03	-0.07	-0.06	0.06	-0.05	-0.03	-0.07
pvIPC7	0	0.11	-0.03	-0.06	-0.08	-0.16	-0.09	-0.08	-0.11	-0.15	-0.16	-0.09	-0.08	-0.11	-0.15
nvIPC7	0.08	0.03	0.07	-0.04	-0.06	-0.03	-0.01	0.02	-0.1	0.04	-0.03	-0.01	0.02	-0.1	0.04
pemIPC7	-0.08	-0.03	-0.03	-0.01	-0.02	-0.13	0.01	-0.27**	-0.09	-0.09	-0.13	0.01	-0.28**	-0.09	-0.09
nemIPC7	0.11	0	0.13	0.04	0.08	-0.06	-0.05	-0.07	-0.07	0.02	-0.06	-0.05	-0.07	-0.07	0.02
conIPC7	0.1	0.01	-0.07	-0.03	-0.09	-0.13	0.1	-0.14	-0.12	-0.1	-0.13	0.1	-0.14	-0.12	-0.1
agreeIPC7	0.02	0.04	-0.05	-0.04	0.03	0.1	0.2*	0.01	0.12	0.02	0.1	0.2*	0.01	0.12	0.02
cnvIPC7	0.1	0.05	-0.01	0.09	0.01	-0.11	-0.11	-0.19*	-0.01	-0.03	-0.11	-0.11	-0.19*	-0.01	-0.03
ReaRIASEC	0.02	0.02	0.11	-0.04	0.01	-0.06	-0.17*	0.09	0	-0.05	-0.06	-0.17*	0.09	0	-0.05
InvRIASEC	-0.02	0.06	0.27**	-0.16	0.11	-0.02	-0.17*	0.06	0.14	-0.06	-0.02	-0.17*	0.06	0.14	-0.06
ArtRIASEC	0	0.02	0.08	-0.07	-0.05	0.01	0.05	0.14	0.03	-0.06	0.01	0.05	0.14	0.04	-0.06
SocRIASEC	0.04	0.08	-0.05	0.07	0.05	-0.09	-0.01	-0.1	0.01	-0.1	-0.09	-0.01	-0.1	0.01	-0.1
EntRIASEC	0.06	-0.05	-0.12	0.07	0	-0.19*	0.07	-0.26**	-0.17*	-0.16	-0.19*	0.07	-0.27**	-0.17*	-0.16
ConRIASEC	0.04	0.05	-0.06	-0.06	0.05	-0.16	-0.07	-0.14	-0.11	-0.12	-0.16	-0.07	-0.15	-0.11	-0.12

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 10 – *cont'd.*

	FisherG	FisherG1	FisherG2	FisherG3	FisherG4	FisherCstat	FisherCstat1	FisherCstat2	FisherCstat3	FisherCstat4	Fisherrz	Fisherrz1	Fisherrz2	Fisherrz3	Fisherrz4
numpredout	0.04	0.03	0.04	0.15	0.09	-0.09	-0.1	-0.02	0	-0.06	-0.09	-0.1	-0.02	0	-0.06
numratees	0.12	0.14	0.08	0.19*	-0.06	0.06	0.04	0.06	0.08	0.07	0.06	0.04	0.05	0.08	0.07
lengthtimepred	0.02	0	-0.03	0.1	0.05	-0.09	-0.12	-0.06	0.08	-0.1	-0.09	-0.12	-0.06	0.08	-0.1
numpredoutformal	-0.16	-0.03	0.01	-0.07	-0.13	0.05	0.08	0.12	0.04	-0.06	0.05	0.08	0.12	0.04	-0.06
numpredouttraining	-0.17*	-0.21*	0	0.05	-0.09	0.01	0.17*	0.05	-0.06	-0.08	0.01	0.17*	0.04	-0.06	-0.08
numpredouttrainingformal	-0.22**	-0.22**	-0.09	-0.13	-0.14	0.03	0.1	-0.03	-0.06	-0.04	0.03	0.1	-0.03	-0.06	-0.04
lengthtraining	-0.12	-0.09	-0.11	-0.05	-0.08	-0.1	-0.02	-0.12	-0.1	-0.07	-0.1	-0.02	-0.12	-0.1	-0.07
amtstats	0.09	0.01	-0.07	-0.06	0.04	0.03	0.07	0.05	-0.07	0	0.03	0.07	0.05	-0.07	0
amtdecisionmkg	-0.03	0.03	-0.17*	0.01	-0.09	-0.13	0	-0.19*	-0.11	-0.11	-0.13	0	-0.19*	-0.11	-0.11
TrPerExprAware	-0.19*	-0.16	0.02	-0.1	-0.14	0.23**	0.23**	0.05	0.03	0.13	0.23**	0.23**	0.05	0.03	0.13
TrPerSaidUsed	-0.2*	-0.16	0.02	-0.11	-0.11	0.24**	0.18*	0.07	0.09	0.16	0.24**	0.18*	0.07	0.09	0.16
TrPerUsedCorrectly	-0.03	0.04	0.09	-0.07	0.02	0.16	0.11	0.06	0.15	0.09	0.16	0.11	0.06	0.15	0.09
TrPerSaidUsed3rdVar	-0.02	0.03	-0.06	-0.02	-0.08	0.09	0.12	0.12	0.1	0.02	0.09	0.12	0.13	0.1	0.02

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 10 – *cont'd.*

	Cwtx1x2_s	Cwtx1x2_1_s	Cwtx1x2_2_s	Cwtx1x2_3_s	Cwtx1x2_4_s	ConfAbsOverallPre	ConfAbs1	ConfAbs2	ConfAbs3	ConfAbs4	ConfAbsOverallPost	ConfRelOverallPre	ConfRel1	ConfRel2	ConfRel3	ConfRel4	ConfRelOverallPost
Fisherra	0.31**	0.31**	0.19*	0.17*	0.32**	-0.15	-0.08	0.01	-0.06	-0.1	0	-0.06	-0.11	-0.13	-0.13	-0.13	-0.06
Fisherra1	0.17*	0.13	0.15	-0.04	-0.03	-0.23*	-0.2*	-0.01	0.05	-0.01	0.03	-0.1	-0.14	-0.12	-0.02	-0.02	-0.04
Fisherra2	0.21*	0.06	0.27**	-0.06	0.14	0	0.03	0.1	0	0.05	0.07	0.11	0.02	-0.06	-0.02	0.04	0.02
Fisherra3	0.2*	0.25**	0.09	0.31**	0.25**	-0.14	-0.03	-0.09	-0.11	-0.21*	-0.08	-0.16	-0.07	-0.15	-0.15	-0.25**	-0.17*
Fisherra4	0.24**	0.32**	0.07	0.3**	0.42**	-0.04	-0.02	0.03	-0.07	-0.12	-0.05	-0.04	-0.08	-0.02	-0.16	-0.16	-0.03
SS	0.2*	0.13	0.06	0.01	0.12	-0.14	-0.17*	-0.22*	-0.22*	-0.27**	-0.23**	-0.03	-0.21*	-0.24**	-0.19*	-0.24**	-0.13
SS1	0.23**	-0.04	0.11	-0.08	0.01	-0.14	-0.19*	-0.09	-0.01	-0.03	0	-0.03	-0.21*	-0.18*	-0.09	-0.05	-0.02
SS2	0.09	0.1	-0.1	-0.07	0.03	-0.02	-0.03	-0.18*	-0.19*	-0.18*	-0.19*	0.09	-0.04	-0.16	-0.12	-0.14	-0.06
SS3	0.15	0.17*	0.15	0.08	0.2*	-0.17	-0.17*	-0.19*	-0.17	-0.2*	-0.19*	-0.11	-0.19*	-0.22**	-0.16	-0.2*	-0.15
SS4	0.14	0.14	0.03	0.05	0.08	-0.07	-0.08	-0.15	-0.23**	-0.31**	-0.22*	-0.01	-0.14	-0.12	-0.15	-0.24**	-0.13
FisherRslin	0.79**	0	0.39**	0.37**	0.12	0.15	0.09	0.21*	0.23**	0.13	0.2*	0.08	0.16	0.26**	0.2*	0.17	0.14
FisherRslin1	0.36**	0.12	0.13	0.18*	0.14	0.24*	0.17*	0.04	0	-0.03	-0.02	0.04	0.12	0.11	0.02	0.03	0
FisherRslin2	0.35**	0.06	0.4**	0.21*	0	0.04	0.02	0.16	0.07	0.08	0.11	0.04	0.07	0.16	0.06	0.13	0
FisherRslin3	0.3**	-0.1	0.16	0.4**	0.1	0.03	0.02	0.03	0.07	0.01	0.05	-0.02	-0.01	0.04	0.04	0.02	0.09
FisherRslin4	0.19*	-0.1	0.04	0.13	0.13	0.03	-0.05	-0.01	0.06	0.1	0.01	-0.02	0.03	0.07	-0.01	0.12	0.06
FisherG	0.54**	-0.07	0.08	0.1	-0.04	-0.01	-0.1	-0.04	0	-0.04	0.04	-0.03	-0.03	0.02	-0.03	0.03	0.01
FisherG1	0.48**	-0.13	0.18*	0.08	0.01	-0.04	-0.1	-0.06	-0.05	-0.07	-0.11	-0.03	-0.06	-0.05	-0.13	-0.07	-0.12
FisherG2	0.35**	-0.04	0.11	-0.03	0.12	-0.15	-0.18*	0	0.03	-0.02	-0.02	-0.11	-0.07	0	0.01	0.02	-0.04
FisherG3	0.41**	-0.05	0.18*	0.15	-0.01	-0.11	-0.01	0.06	-0.01	-0.05	0.07	-0.12	-0.05	-0.05	-0.06	-0.08	-0.04
FisherG4	0.63**	-0.05	0.14	0.24**	0.15	-0.01	-0.15	-0.02	0.02	-0.12	-0.03	0.05	-0.1	0.06	-0.04	-0.06	0.04

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 10 – *cont'd.*

	Cwtx1x2_s	Cwtx1x2_1_s	Cwtx1x2_2_s	Cwtx1x2_3_s	Cwtx1x2_4_s	ConfAbsOverallPre	ConfAbs1	ConfAbs2	ConfAbs3	ConfAbs4	ConfAbsOverallPost	ConfRelOverallPre	ConfRel1	ConfRel2	ConfRel3	ConfRel4	ConfRelOverallPost
FisherCstat	0.03	0.32**	0.12	0.13	0.28**	-0.12	-0.03	-0.01	-0.1	-0.09	-0.08	-0.05	-0.08	-0.12	-0.11	-0.11	-0.1
FisherCstat1	-0.07	0.15	0.07	0.01	-0.02	-0.15	-0.19*	0.01	0.01	0.01	-0.04	-0.08	-0.13	-0.12	0.01	-0.02	-0.08
FisherCstat2	0.09	0.09	0.21*	-0.02	0.14	0.02	0.08	0.04	-0.01	0.06	0.03	0.11	0.04	-0.06	-0.01	0.04	0.01
FisherCstat3	-0.01	0.19*	0.09	0.18*	0.18*	-0.14	-0.01	-0.05	-0.08	-0.18*	-0.1	-0.16	-0.09	-0.09	-0.12	-0.21*	-0.18*
FisherCstat4	0.01	0.29**	-0.02	0.15	0.27**	-0.03	0	-0.02	-0.15	-0.15	-0.1	-0.04	-0.08	-0.05	-0.17	-0.18*	-0.08
Fisherrz	0.03	0.32**	0.12	0.13	0.28**	-0.12	-0.03	-0.01	-0.1	-0.09	-0.08	-0.05	-0.08	-0.12	-0.11	-0.11	-0.1
Fisherrz1	-0.07	0.15	0.07	0.01	-0.02	-0.15	-0.19*	0.01	0.01	0.01	-0.04	-0.08	-0.13	-0.12	0.01	-0.02	-0.08
Fisherrz2	0.09	0.09	0.21*	-0.02	0.13	0.02	0.08	0.04	-0.01	0.06	0.03	0.11	0.04	-0.06	-0.01	0.04	0.01
Fisherrz3	-0.01	0.19*	0.09	0.18*	0.18*	-0.14	-0.01	-0.05	-0.08	-0.18*	-0.1	-0.16	-0.09	-0.09	-0.12	-0.21*	-0.18*
Fisherrz4	0.01	0.29**	-0.02	0.15	0.27**	-0.03	0	-0.02	-0.15	-0.15	-0.1	-0.04	-0.08	-0.05	-0.17	-0.18*	-0.08
Cwtx1x2_s	1**	0.19*	0.49**	0.44**	0.35**	0.06	0.03	0.16	0.21*	0.08	0.17	0.06	0.1	0.17	0.16	0.15	0.11
Cwtx1x2_1_s	0.19*	1**	0.24**	0.3**	0.41**	0.01	0.15	0.05	0.02	0.1	0.06	0.09	0.14	0.01	0.09	0.16	-0.02
Cwtx1x2_2_s	0.49**	0.24**	1**	0.39**	0.38**	0.17	0.04	0.2*	0.19*	0.13	0.18*	0.15	0.04	0.13	0.11	0.16	0.07
Cwtx1x2_3_s	0.44**	0.3**	0.39**	1**	0.51**	0.1	0.16	0.15	0.19*	0.06	0.12	-0.03	0.1	0.11	0.15	0.05	0.05
Cwtx1x2_4_s	0.35**	0.41**	0.38**	0.51**	1**	0.09	0.11	0.03	0.16	0.11	0.04	0.05	0.07	0.03	0.07	0.13	0.06
ConfAbsOverallPre	0.06	0.01	0.17	0.1	0.09	1**	0.55**	0.51**	0.48**	0.48**	0.53**	0.76**	0.58**	0.57**	0.4**	0.54**	0.56**
ConfAbs1	0.03	0.15	0.04	0.16	0.11	0.55**	1**	0.59**	0.54**	0.55**	0.61**	0.41**	0.73**	0.5**	0.5**	0.53**	0.52**
ConfAbs2	0.16	0.05	0.2*	0.15	0.03	0.51**	0.59**	1**	0.75**	0.67**	0.73**	0.47**	0.6**	0.71**	0.63**	0.61**	0.63**
ConfAbs3	0.21*	0.02	0.19*	0.19*	0.16	0.48**	0.54**	0.75**	1**	0.78**	0.73**	0.43**	0.58**	0.66**	0.79**	0.74**	0.67**
ConfAbs4	0.08	0.1	0.13	0.06	0.11	0.48**	0.55**	0.67**	0.78**	1**	0.71**	0.43**	0.59**	0.6**	0.68**	0.85**	0.68**

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 10 – *cont'd.*

	Cwtx1x2_s	Cwtx1x2_1_s	Cwtx1x2_2_s	Cwtx1x2_3_s	Cwtx1x2_4_s	ConfAbsOverallPre	ConfAbs1	ConfAbs2	ConfAbs3	ConfAbs4	ConfAbsOverallPost	ConfRelOverallPre	ConfRel1	ConfRel2	ConfRel3	ConfRel4	ConfRelOverallPost
ConfAbsOverallPost	0.17	0.06	0.18*	0.12	0.04	0.53**	0.61**	0.73**	0.73**	0.71**	1**	0.41**	0.6**	0.68**	0.65**	0.71**	0.8**
ConfRelOverallPre	0.06	0.09	0.15	-0.03	0.05	0.76**	0.41**	0.47**	0.43**	0.43**	0.41**	1**	0.56**	0.59**	0.47**	0.57**	0.56**
ConfRel1	0.1	0.14	0.04	0.1	0.07	0.58**	0.73**	0.6**	0.58**	0.59**	0.6**	0.56**	1**	0.74**	0.67**	0.69**	0.64**
ConfRel2	0.17	0.01	0.13	0.11	0.03	0.57**	0.5**	0.71**	0.66**	0.6**	0.68**	0.59**	0.74**	1**	0.76**	0.7**	0.77**
ConfRel3	0.16	0.09	0.11	0.15	0.07	0.4**	0.5**	0.63**	0.79**	0.68**	0.65**	0.47**	0.67**	0.76**	1**	0.81**	0.72**
ConfRel4	0.15	0.16	0.16	0.05	0.13	0.54**	0.53**	0.61**	0.74**	0.85**	0.71**	0.57**	0.69**	0.7**	0.81**	1**	0.74**
ConfRelOverallPost	0.11	-0.02	0.07	0.05	0.06	0.56**	0.52**	0.63**	0.67**	0.68**	0.8**	0.56**	0.64**	0.77**	0.72**	0.74**	1**
ACT_COMP_SCR	0.17	0.28**	0.01	0.26**	0.47**	0.18	0.06	-0.01	0	0.07	0.06	0.19	0.07	0.06	0.11	0.07	0.2*
ACT_ENGL_SCR	0.14	0.31**	0	0.2*	0.4**	0.14	0	-0.07	-0.03	0.02	0.03	0.16	0.04	-0.03	0.02	0	0.15
ACT_ENGWR_SCR	0.12	0.32**	-0.01	0.18	0.4**	0.13	-0.04	-0.1	-0.06	0.03	-0.02	0.18	0.04	-0.03	0.03	0.01	0.11
ACT_MATH_SCR	0.17	0.18	-0.01	0.29**	0.42**	0.15	0.04	-0.11	0.01	0.03	0.03	0.17	0.07	0.05	0.1	0.04	0.18
ACT_READ_SCR	0.15	0.29**	-0.01	0.17	0.41**	0.19	0.1	0.07	0.03	0.11	0.11	0.18	0.09	0.12	0.16	0.13	0.2*
ACT_SCIRE_SCR	0.14	0.2*	0.03	0.23*	0.42**	0.19	0.07	0.02	0.03	0.08	0.08	0.19	0.09	0.11	0.12	0.08	0.23*
CUM_GPA	0.25**	0.2*	0.1	0.11	0.22**	0.02	-0.16	-0.08	-0.09	-0.09	-0.18*	0.11	-0.08	-0.04	-0.06	-0.07	-0.12
TOT_ACAD_HOURS	0.1	0.07	0.16	0.03	0.02	0.04	-0.12	-0.11	-0.06	-0.11	-0.08	0.12	0.08	0.04	0.1	-0.02	0
HS_RANK_PCT	0.14	0.02	0.22*	-0.12	0.09	0.03	-0.18	-0.1	-0.15	-0.11	-0.14	0.13	-0.1	-0.1	-0.19	-0.16	0.01
Gender	0	0.09	0.03	0.11	0.1	0.21*	0.24**	0.15	0.25**	0.18*	0.17	0.3**	0.27**	0.27**	0.33**	0.23*	0.26**

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 10 – *cont'd.*

	Cwtx1x2_s	Cwtx1x2_1_s	Cwtx1x2_2_s	Cwtx1x2_3_s	Cwtx1x2_4_s	ConfAbsOverallPre	ConfAbs1	ConfAbs2	ConfAbs3	ConfAbs4	ConfAbsOverallPost	ConfRelOverallPre	ConfRel1	ConfRel2	ConfRel3	ConfRel4	ConfRelOverallPost
extroGoldberg	0.04	0.08	0.02	-0.15	0.05	0.22*	0.1	0.12	0.18*	0.21*	0.17	0.3**	0.26**	0.25**	0.24**	0.29**	0.18*
neurotGoldberg	-0.04	-0.15	-0.01	0.01	-0.14	-0.19*	-0.18*	-0.16	-0.16	-0.15	-0.21*	-0.19*	-0.18*	-0.17	-0.22*	-0.22*	-0.17
intelGoldberg	0.03	0.1	0.17*	0.1	0.18*	0.26**	0.2*	0.08	0.11	0.2*	0.16	0.27**	0.19*	0.07	0.07	0.2*	0.12
agreeGoldberg	0.13	0.05	0.05	0.1	0.09	-0.07	-0.03	-0.04	0.01	0.03	-0.04	-0.09	-0.03	-0.06	-0.03	0.01	-0.19*
conGoldberg	0.14	0.03	0.08	0.11	-0.09	0.14	0.13	0.11	0.2*	0.13	0.13	0.16	0.15	0.19*	0.24**	0.17	0.13
pvIPC7	0.03	-0.12	0.03	-0.1	0	0.38**	0.26**	0.17*	0.35**	0.34**	0.28**	0.35**	0.33**	0.3**	0.32**	0.37**	0.26**
nvIPC7	-0.15	-0.1	-0.12	-0.24**	-0.26**	0.06	0.01	-0.04	-0.07	-0.06	0.06	0.03	-0.01	0	-0.04	-0.04	0.12
pemIPC7	0.03	0.12	0.01	-0.11	0.09	0.17	0.07	0.13	0.2*	0.21*	0.17*	0.16	0.17	0.18*	0.21*	0.27**	0.09
nemIPC7	-0.05	-0.18*	-0.02	0.08	-0.16	-0.18	-0.23**	-0.19*	-0.28**	-0.28**	-0.28**	-0.22*	-0.19*	-0.16	-0.31**	-0.33**	-0.29**
conIPC7	-0.02	-0.11	-0.1	-0.1	-0.28**	0.05	0.1	0.07	0.14	0.07	0.08	0.06	0.06	0.1	0.18*	0.13	0.08
agreeIPC7	-0.03	0.07	-0.06	0.05	0.07	-0.19*	-0.12	-0.06	-0.11	-0.12	-0.07	-0.23*	-0.22*	-0.16	-0.15	-0.16	-0.2*
cnvIPC7	0.06	-0.11	-0.11	0	-0.04	0.02	0.05	0.05	0.06	0.05	0.06	-0.09	0.09	0.14	0.1	0.1	0.07
ReaRIASEC	-0.02	-0.05	-0.01	0	0.04	-0.03	0.1	-0.04	0.04	-0.09	-0.03	-0.03	0	-0.02	0.06	-0.07	0.13
InvRIASEC	0.1	-0.03	0.07	0.01	0.15	-0.1	0.04	-0.04	0.01	-0.09	-0.05	0.03	0.01	0.03	-0.02	-0.07	0.1
ArtRIASEC	-0.07	0	0.03	-0.03	0.13	-0.17	-0.01	-0.25**	-0.21*	-0.17	-0.17	-0.16	-0.1	-0.21*	-0.18*	-0.21*	-0.15
SocRIASEC	0.08	-0.08	-0.09	-0.09	0.02	-0.16	-0.02	-0.13	-0.05	-0.12	-0.14	-0.19*	-0.1	-0.16	-0.05	-0.12	-0.16
EntRIASEC	-0.07	-0.1	-0.08	-0.15	-0.16	0.07	0.07	0.09	0.18*	0.09	0.17	0.03	0.01	0.08	0.09	0.11	0.18*
ConRIASEC	0.04	-0.07	0.04	-0.02	-0.07	0.05	0.08	-0.03	0.06	-0.02	0.09	0.08	-0.02	0.01	0.05	0.01	0.2*

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 10 – *cont'd.*

	Cwtx1x2_s	Cwtx1x2_1_s	Cwtx1x2_2_s	Cwtx1x2_3_s	Cwtx1x2_4_s	ConfAbsOverallPre	ConfAbs1	ConfAbs2	ConfAbs3	ConfAbs4	ConfAbsOverallPost	ConfRelOverallPre	ConfRel1	ConfRel2	ConfRel3	ConfRel4	ConfRelOverallPost
numpredout	0.07	-0.04	0.07	-0.05	0.02	0.21*	0.04	0.12	0.1	0.04	0.07	0.24*	0.04	0.12	0.05	-0.03	0.07
numratees	0.06	0.08	0.08	0.04	0	0.13	-0.03	-0.02	-0.03	-0.05	0	0.18	0.06	0.05	0.01	-0.04	0.02
lengthtimepred	0.04	-0.09	0	0.01	0	0.13	0.13	0.1	0.13	0.1	0.12	0.17	0.05	0.04	0.04	0	0.07
numpredoutformal	-0.05	-0.01	-0.02	-0.14	-0.08	0.07	-0.01	-0.03	0.01	0.03	-0.08	0.14	0.07	0.04	0.08	0.01	0.02
numpredouttraining	-0.01	0.02	0.05	0.01	0.09	0.19*	0.08	0.08	0.14	0.08	0.07	0.21*	0.05	0.03	0.13	0.04	0.1
numpredouttrainingformal	-0.05	-0.02	0.06	0.02	0.04	0.2*	0.09	0.06	0.05	0.06	0.1	0.13	0.06	0.07	0.06	0.01	0.09
lengthtraining	-0.11	-0.13	-0.19*	-0.18*	-0.05	-0.08	-0.08	-0.05	-0.01	0.02	-0.07	-0.04	-0.02	-0.03	0.03	-0.06	0.01
amtstats	0.06	0.03	0.06	0.03	0.02	0.1	0.02	-0.06	0.03	-0.01	0.04	0.12	0.1	0.02	0.06	0.01	0.11
amtdecisionmkg	-0.04	0.01	-0.02	-0.01	0.04	-0.02	0.08	0.02	0.07	0.16	-0.03	0.12	0.07	0.07	0.09	0.11	0.03
TrPerExprAware	-0.07	0.03	0.04	-0.07	0.18*	0	0.05	-0.04	0.01	0.02	-0.04	0.08	0.04	-0.04	0.03	-0.01	-0.04
TrPerSaidUsed	-0.11	0.11	-0.05	-0.06	0.2*	-0.04	0.06	-0.05	-0.03	0.01	-0.1	0.04	0.05	-0.07	0.02	-0.04	-0.09
TrPerUsedCorrectly	0.13	0.13	0.17	0.19*	0.35**	-0.07	-0.08	-0.05	-0.05	-0.05	-0.13	-0.05	-0.12	-0.14	-0.04	-0.12	-0.15
TrPerSaidUsed3rdVar	-0.02	0.09	0.04	-0.06	-0.13	0.14	0.06	0.02	0.15	0.18	0.18	0.11	0.07	0.13	0.1	0.13	0.19*

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 10 – *cont'd.*

	ACT_COMP_SCR	ACT_ENGL_SCR	ACT_ENGWR_SCR	ACT_MATH_SCR	ACT_READ_SCR	ACT_SCIRE_SCR	CUM_GPA	TOT_ACAD_HOURS	HS_RANK_PCT	Gender
Fisherra	0.32**	0.38**	0.39**	0.2*	0.28**	0.24*	0.24**	0.09	0.2*	0.02
Fisherra1	-0.07	0.07	0.09	-0.17	-0.02	-0.18	0.04	0.11	0.17	-0.16
Fisherra2	0.27**	0.33**	0.34**	0.18	0.22*	0.27**	0.11	0.15	0.3**	0
Fisherra3	0.28**	0.27**	0.29**	0.24*	0.21*	0.21*	0.18*	0.05	-0.06	0.07
Fisherra4	0.36**	0.34**	0.3**	0.28**	0.32**	0.34**	0.23**	-0.05	0.11	0.1
SS	0.25**	0.23*	0.26**	0.22*	0.22*	0.22*	0.11	0.12	0.1	0.05
SS1	-0.05	0.01	-0.01	-0.08	-0.05	-0.04	-0.05	0.04	0.08	-0.01
SS2	0.27**	0.23*	0.25*	0.23*	0.25**	0.26**	0.11	0.12	0.05	0.08
SS3	0.22*	0.22*	0.26**	0.23*	0.14	0.12	0.1	0.13	0.05	-0.02
SS4	0.3**	0.23*	0.26**	0.25**	0.3**	0.26**	0.18*	0.06	0.08	0.08
FisherRslin	-0.1	-0.09	-0.14	-0.04	-0.14	-0.06	0.07	0.05	0.03	-0.02
FisherRslin1	0.11	0.11	0.06	0.13	0.05	0.08	0.09	0.01	0.04	0.02
FisherRslin2	-0.08	-0.18	-0.23*	-0.15	-0.04	-0.17	0	-0.06	0.02	-0.14
FisherRslin3	-0.13	-0.15	-0.23*	0	-0.18	-0.13	-0.07	-0.12	-0.06	-0.1
FisherRslin4	-0.09	-0.08	-0.13	0.02	-0.17	-0.08	-0.09	-0.07	0.01	-0.08
FisherG	0.07	0.09	0.12	0.07	0.08	-0.02	0.19*	-0.01	0.09	-0.07
FisherG1	-0.15	-0.11	-0.09	-0.12	-0.2*	-0.14	-0.02	0	0.12	0
FisherG2	0.24*	0.19*	0.24*	0.17	0.22*	0.25**	0.12	0.03	0.09	0.01
FisherG3	-0.07	-0.05	-0.09	-0.09	-0.06	-0.16	0.04	-0.02	-0.05	-0.09
FisherG4	0.12	0.05	0.01	0.15	0.12	0.1	0.25**	-0.06	0.13	-0.03

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 10 – *cont'd.*

	ACT_COMP_SCR	ACT_ENGL_SCR	ACT_ENGWR_SCR	ACT_MATH_SCR	ACT_READ_SCR	ACT_SCIRE_SCR	CUM_GPA	TOT_ACAD_HOURS	HS_RANK_PCT	Gender
FisherCstat	0.26**	0.31**	0.33**	0.13	0.23*	0.22*	0.13	0.11	0.17	0.07
FisherCstat1	-0.03	0.1	0.13	-0.16	0.01	-0.14	0.04	0.15	0.12	-0.12
FisherCstat2	0.2*	0.27**	0.26**	0.14	0.12	0.2*	0.08	0.13	0.25*	0
FisherCstat3	0.16	0.15	0.18	0.09	0.16	0.11	0.09	0.05	-0.09	0.1
FisherCstat4	0.27**	0.28**	0.25*	0.2*	0.22*	0.24*	0.14	-0.05	0.08	0.13
Fisherrz	0.26**	0.31**	0.33**	0.13	0.23*	0.22*	0.13	0.11	0.17	0.07
Fisherrz1	-0.03	0.1	0.13	-0.16	0.01	-0.14	0.04	0.15	0.12	-0.12
Fisherrz2	0.2*	0.27**	0.26**	0.14	0.12	0.2*	0.08	0.13	0.25*	0
Fisherrz3	0.16	0.15	0.18	0.09	0.16	0.11	0.09	0.05	-0.09	0.1
Fisherrz4	0.27**	0.28**	0.25*	0.2*	0.22*	0.24*	0.14	-0.05	0.08	0.13
Cwtx1x2_s	0.17	0.14	0.12	0.17	0.15	0.14	0.25**	0.1	0.14	0
Cwtx1x2_1_s	0.28**	0.31**	0.32**	0.18	0.29**	0.2*	0.2*	0.07	0.02	0.09
Cwtx1x2_2_s	0.01	0	-0.01	-0.01	-0.01	0.03	0.1	0.16	0.22*	0.03
Cwtx1x2_3_s	0.26**	0.2*	0.18	0.29**	0.17	0.23*	0.11	0.03	-0.12	0.11
Cwtx1x2_4_s	0.47**	0.4**	0.4**	0.42**	0.41**	0.42**	0.22**	0.02	0.09	0.1
ConfAbsOverallPre	0.18	0.14	0.13	0.15	0.19	0.19	0.02	0.04	0.03	0.21*
ConfAbs1	0.06	0	-0.04	0.04	0.1	0.07	-0.16	-0.12	-0.18	0.24**
ConfAbs2	-0.01	-0.07	-0.1	-0.11	0.07	0.02	-0.08	-0.11	-0.1	0.15
ConfAbs3	0	-0.03	-0.06	0.01	0.03	0.03	-0.09	-0.06	-0.15	0.25**
ConfAbs4	0.07	0.02	0.03	0.03	0.11	0.08	-0.09	-0.11	-0.11	0.18*
ConfAbsOverallPost	0.06	0.03	-0.02	0.03	0.11	0.08	-0.18*	-0.08	-0.14	0.17

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 10 – *cont'd.*

	ACT_COMP_SCR	ACT_ENGL_SCR	ACT_ENGWR_SCR	ACT_MATH_SCR	ACT_READ_SCR	ACT_SCIRE_SCR	CUM_GPA	TOT_ACAD_HOURS	HS_RANK_PCT	Gender
ConfRelOverallPre	0.19	0.16	0.18	0.17	0.18	0.19	0.11	0.12	0.13	0.3**
ConfRel1	0.07	0.04	0.04	0.07	0.09	0.09	-0.08	0.08	-0.1	0.27**
ConfRel2	0.06	-0.03	-0.03	0.05	0.12	0.11	-0.04	0.04	-0.1	0.27**
ConfRel3	0.11	0.02	0.03	0.1	0.16	0.12	-0.06	0.1	-0.19	0.33**
ConfRel4	0.07	0	0.01	0.04	0.13	0.08	-0.07	-0.02	-0.16	0.23*
ConfRelOverallPost	0.2*	0.15	0.11	0.18	0.2*	0.23*	-0.12	0	0.01	0.26**
ACT_COMP_SCR	1**	0.87**	0.82**	0.87**	0.87**	0.87**	0.45**	0.07	0.32**	0.09
ACT_ENGL_SCR	0.87**	1**	0.95**	0.72**	0.7**	0.69**	0.47**	0.12	0.37**	-0.06
ACT_ENGWR_SCR	0.82**	0.95**	1**	0.67**	0.68**	0.64**	0.47**	0.32**	0.3**	-0.06
ACT_MATH_SCR	0.87**	0.72**	0.67**	1**	0.63**	0.77**	0.43**	0.01	0.31**	0.17
ACT_READ_SCR	0.87**	0.7**	0.68**	0.63**	1**	0.71**	0.43**	0.1	0.2	0.05
ACT_SCIRE_SCR	0.87**	0.69**	0.64**	0.77**	0.71**	1**	0.28**	0.07	0.35**	0.19*
CUM_GPA	0.45**	0.47**	0.47**	0.43**	0.43**	0.28**	1**	0.11	0.24*	-0.11
TOT_ACAD_HOURS	0.07	0.12	0.32**	0.01	0.1	0.07	0.11	1**	0.12	0.09
HS_RANK_PCT	0.32**	0.37**	0.3**	0.31**	0.2	0.35**	0.24*	0.12	1**	0.02
Gender	0.09	-0.06	-0.06	0.17	0.05	0.19*	-0.11	0.09	0.02	1**

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 10 – *cont'd.*

	ACT_COMP_SCR	ACT_ENGL_SCR	ACT_ENGWR_SCR	ACT_MATH_SCR	ACT_READ_SCR	ACT_SCIRE_SCR	CUM_GPA	TOT_ACAD_HOURS	HS_RANK_PCT	Gender
extroGoldberg	-0.09	-0.11	-0.12	-0.08	-0.03	-0.06	0.07	-0.01	0.01	0.03
neurotGoldberg	-0.15	-0.15	-0.19	-0.12	-0.16	-0.09	-0.03	-0.07	-0.05	-0.2*
intelGoldberg	0.25**	0.26**	0.26**	0.2*	0.27**	0.1	0.14	0.12	0.18	-0.06
agreeGoldberg	-0.23*	-0.17	-0.16	-0.2*	-0.21*	-0.29**	0.04	-0.04	-0.12	-0.16
conGoldberg	-0.06	-0.09	-0.09	-0.03	0	-0.13	0.11	0.05	-0.02	-0.01
pvIPC7	-0.09	-0.11	-0.09	-0.02	-0.05	-0.03	0.02	-0.04	-0.09	0.23**
nvIPC7	-0.01	-0.01	-0.05	0.02	-0.02	0.02	-0.21*	-0.03	0.18	0.1
pemIPC7	-0.1	-0.09	-0.07	-0.12	-0.02	-0.08	0.08	-0.04	-0.05	-0.12
nemIPC7	-0.12	-0.11	-0.13	-0.08	-0.1	-0.11	0.1	-0.07	-0.01	-0.19*
conIPC7	-0.17	-0.15	-0.17	-0.13	-0.11	-0.27**	0.09	-0.01	-0.13	-0.1
agreeIPC7	-0.01	0.1	0.11	-0.12	0.02	-0.15	-0.06	-0.02	-0.12	-0.25**
cnvIPC7	-0.21*	-0.29**	-0.27**	-0.09	-0.18	-0.21*	-0.04	-0.13	-0.07	-0.01
ReaRIASEC	0.18	0.08	0	0.27**	0.09	0.24*	-0.12	0.03	0.07	0.37**
InvRIASEC	0.29**	0.22*	0.22*	0.35**	0.17	0.33**	0.06	-0.02	0.18	0.15
ArtRIASEC	0.22*	0.28**	0.26**	0.16	0.17	0.09	0.03	0.04	0.12	-0.24**
SocRIASEC	-0.16	-0.05	-0.08	-0.2*	-0.14	-0.2*	-0.05	-0.04	-0.24*	-0.31**
EntRIASEC	-0.19*	-0.16	-0.2*	-0.13	-0.14	-0.26**	-0.06	-0.19*	-0.23*	-0.13
ConRIASEC	0.08	0.05	-0.1	0.09	0.04	0.14	-0.08	-0.01	0.11	0.08

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 10 – *cont'd.*

	ACT_COMP_SCR	ACT_ENGL_SCR	ACT_ENGWR_SCR	ACT_MATH_SCR	ACT_READ_SCR	ACT_SCIRE_SCR	CUM_GPA	TOT_ACAD_HOURS	HS_RANK_PCT	Gender
numpredout	-0.07	-0.09	0	-0.05	-0.07	0.04	-0.07	0.16	0.02	0.01
numratees	-0.07	0	0.04	-0.1	-0.1	0.06	0.04	0.25**	0.09	0.02
lengthtimepred	-0.04	-0.06	0	-0.03	-0.02	0.04	0.01	0.04	0.1	0.01
numpredoutformal	-0.16	-0.18	-0.06	-0.14	-0.13	-0.08	-0.01	0.31**	-0.08	0.19*
numpredouttraining	0.08	0.11	0.21*	0.01	0.09	0.13	-0.06	0.21*	0.02	-0.05
numpredouttrainingformal	-0.05	-0.02	0.06	-0.09	-0.08	0.05	-0.07	0.16	-0.09	0.04
lengthtraining	-0.1	-0.14	-0.07	-0.05	-0.1	0	-0.06	0.08	-0.01	0.11
amtstats	-0.06	0.03	0.1	-0.14	-0.01	-0.04	0	0.4**	0	-0.05
amtdecisionmkg	-0.19*	-0.17	-0.07	-0.21*	-0.13	-0.19	0.01	0.11	-0.19	0.13
TrPerExprAware	0.02	0.04	0.09	-0.03	0.1	0.03	0.04	0.11	0	-0.02
TrPerSaidUsed	0.17	0.13	0.17	0.08	0.25**	0.18	0.13	0.13	0.06	0.02
TrPerUsedCorrectly	0.23*	0.21*	0.27**	0.16	0.23*	0.15	0.22*	0.05	0.06	0
TrPerSaidUsed3rdVar	-0.01	0.02	0.03	0.03	-0.06	0	-0.09	0.07	-0.14	0.09

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 10 – *cont'd.*

	extroGoldberg	neurotGoldberg	intelGoldberg	agreeGoldberg	conGoldberg	pvIPC7	nvIPC7	pemIPC7	nemIPC7	conIPC7	agreeIPC7	cnvIPC7	ReaRIASEC	InvRIASEC	ArtRIASEC	SocRIASEC	EntRIASEC	ConRIASEC
Fisherra	-0.16	-0.13	0.07	0.04	-0.06	-0.14	-0.04	-0.13	-0.07	-0.17*	0.13	-0.14	-0.06	0.03	0.01	-0.11	-0.18*	0.2*
Fisherra1	-0.06	-0.05	0.1	0.14	0.04	-0.07	0.04	0	-0.04	0.11	0.23**	-0.13	-0.14	-0.12	0	-0.02	0.07	-0.06
Fisherra2	-0.23**	-0.03	0.18*	-0.1	-0.06	-0.02	0.01	-0.24**	-0.04	-0.14	-0.01	-0.17*	0.03	0.18*	0.11	-0.13	-0.23**	-0.14
Fisherra3	-0.19*	-0.13	-0.08	0.05	-0.03	-0.13	-0.09	-0.15	-0.03	-0.15	0.07	-0.03	0.02	0.08	-0.01	-0.05	-0.16	-0.15
Fisherra4	-0.05	-0.09	0.01	0.03	-0.06	-0.14	-0.08	-0.04	-0.03	-0.2*	0.07	-0.05	-0.04	0.02	-0.03	-0.11	-0.15	-0.14
SS	-0.12	-0.06	-0.03	-0.11	-0.08	-0.18*	0.01	-0.14	-0.05	-0.19*	0.04	-0.22**	0.18*	0.17*	0.09	0.07	-0.08	-0.05
SS1	-0.04	0.03	-0.01	-0.07	-0.04	0	0.1	-0.05	-0.05	-0.07	0.04	-0.14	0.08	0.02	0.01	0.11	0.14	0.09
SS2	-0.11	-0.05	-0.1	-0.16	-0.06	-0.11	0.01	-0.15	-0.07	-0.07	-0.01	-0.12	0.17*	0.16	0.02	0.02	-0.08	0
SS3	-0.08	-0.06	-0.01	-0.04	-0.07	-0.24**	-0.04	-0.06	-0.04	-0.22**	0.04	-0.24**	0.17*	0.16	0.18*	0.07	-0.07	-0.11
SS4	-0.1	-0.11	0.03	-0.05	-0.04	-0.17*	-0.03	-0.13	-0.03	-0.14	0.04	-0.15	0.1	0.16	0.05	0.02	-0.18*	-0.13
FisherRslin	0.07	0.02	-0.04	0.13	0.15	0.05	-0.15	0.04	0	0.04	-0.04	0.14	-0.02	0.07	-0.12	0.15	-0.03	0.14
FisherRslin1	0.12	-0.08	-0.08	-0.02	0.02	0.01	-0.07	0.09	-0.05	-0.02	0.02	0.12	0.09	0.17*	0.05	0.12	-0.08	0.14
FisherRslin2	0.16*	-0.07	0.1	0.14	0.11	-0.07	-0.06	0.12	-0.02	-0.09	0.09	-0.06	-0.04	-0.1	0.1	0.05	0.04	0.04
FisherRslin3	-0.06	0.12	0.06	0.08	0.13	0.02	-0.14	-0.14	0.08	0.06	-0.01	0.15	0.04	0.07	0.03	0.12	0.06	0.04
FisherRslin4	0	0.06	0.07	0.1	0	0.01	-0.06	0	0.04	-0.05	0.01	0.03	0.14	0.14	0.09	0.11	0.04	0
FisherG	-0.14	0.05	-0.05	-0.05	0.06	0	0.08	-0.08	0.11	0.1	0.02	0.1	0.02	-0.02	0	0.04	0.06	0.04
FisherG1	-0.02	0.02	-0.08	0.08	-0.02	0.11	0.03	-0.03	0	0.01	0.04	0.05	0.02	0.06	0.02	0.08	-0.05	0.05
FisherG2	-0.06	0.06	-0.04	-0.14	-0.09	-0.03	0.07	-0.03	0.13	-0.07	-0.05	-0.01	0.11	0.27**	0.08	-0.05	-0.12	-0.06
FisherG3	-0.01	-0.02	-0.09	-0.03	-0.02	-0.06	-0.04	-0.01	0.04	-0.03	-0.04	0.09	-0.04	-0.16	-0.07	0.07	0.07	-0.06
FisherG4	-0.06	0.03	-0.06	0.06	0.05	-0.08	-0.06	-0.02	0.08	-0.09	0.03	0.01	0.01	0.11	-0.05	0.05	0	0.05

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 10 – *cont'd.*

	extroGoldberg	neurotGoldberg	intelGoldberg	agreeGoldberg	conGoldberg	pvIPC7	nvIPC7	pemIPC7	nemIPC7	conIPC7	agreeIPC7	cnvIPC7	ReaRIASEC	InvRIASEC	ArtRIASEC	SocRIASEC	EntRIASEC	ConRIASEC
FisherCstat	-0.15	-0.08	0.11	0.04	-0.06	-0.16	-0.03	-0.13	-0.06	-0.13	0.1	-0.11	-0.06	-0.02	0.01	-0.09	-0.19*	-0.16
FisherCstat1	-0.03	-0.08	0.12	0.13	0.06	-0.09	-0.01	0.01	-0.05	0.1	0.2*	-0.11	-0.17*	-0.17*	0.05	-0.01	0.07	-0.07
FisherCstat2	-0.25**	-0.05	0.2*	-0.05	-0.05	-0.08	0.02	-0.27**	-0.07	-0.14	0.01	-0.19*	0.09	0.06	0.14	-0.1	-0.26**	-0.14
FisherCstat3	-0.15	-0.09	-0.06	0.04	-0.03	-0.11	-0.1	-0.09	-0.07	-0.12	0.12	-0.01	0	0.14	0.03	0.01	-0.17*	-0.11
FisherCstat4	-0.08	0	-0.01	-0.01	-0.07	-0.15	0.04	-0.09	0.02	-0.1	0.02	-0.03	-0.05	-0.06	-0.06	-0.1	-0.16	-0.12
Fisherrz	-0.15	-0.08	0.11	0.04	-0.06	-0.16	-0.03	-0.13	-0.06	-0.13	0.1	-0.11	-0.06	-0.02	0.01	-0.09	-0.19*	-0.16
Fisherrz1	-0.03	-0.08	0.12	0.13	0.06	-0.09	-0.01	0.01	-0.05	0.1	0.2*	-0.11	-0.17*	-0.17*	0.05	-0.01	0.07	-0.07
Fisherrz2	-0.25**	-0.05	0.2*	-0.05	-0.05	-0.08	0.02	-0.28**	-0.07	-0.14	0.01	-0.19*	0.09	0.06	0.14	-0.1	-0.27**	-0.15
Fisherrz3	-0.16	-0.09	-0.06	0.04	-0.03	-0.11	-0.1	-0.09	-0.07	-0.12	0.12	-0.01	0	0.14	0.04	0.01	-0.17*	-0.11
Fisherrz4	-0.08	0	-0.01	-0.01	-0.07	-0.15	0.04	-0.09	0.02	-0.1	0.02	-0.03	-0.05	-0.06	-0.06	-0.1	-0.16	-0.12
Cwtx1x2_s	0.04	-0.04	0.03	0.13	0.14	0.03	-0.15	0.03	-0.05	-0.02	-0.03	0.06	-0.02	0.1	-0.07	0.08	-0.07	0.04
Cwtx1x2_1_s	0.08	-0.15	0.1	0.05	0.03	-0.12	-0.1	0.12	-0.18*	-0.11	0.07	-0.11	-0.05	-0.03	0	-0.08	-0.1	-0.07
Cwtx1x2_2_s	0.02	-0.01	0.17*	0.05	0.08	0.03	-0.12	0.01	-0.02	-0.1	-0.06	-0.11	-0.01	0.07	0.03	-0.09	-0.08	0.04
Cwtx1x2_3_s	-0.15	0.01	0.1	0.1	0.11	-0.1	-0.24**	-0.11	0.08	-0.1	0.05	0	0	0.01	-0.03	-0.09	-0.15	-0.02
Cwtx1x2_4_s	0.05	-0.14	0.18*	0.09	-0.09	0	-0.26**	0.09	-0.16	-0.28**	0.07	-0.04	0.04	0.15	0.13	0.02	-0.16	-0.07
ConfAbsOverallPre	0.22*	-0.19*	0.26**	-0.07	0.14	0.38**	0.06	0.17	-0.18	0.05	-0.19*	0.02	-0.03	-0.1	-0.17	-0.16	0.07	0.05
ConfAbs1	0.1	-0.18*	0.2*	-0.03	0.13	0.26**	0.01	0.07	-0.23**	0.1	-0.12	0.05	0.1	0.04	-0.01	-0.02	0.07	0.08
ConfAbs2	0.12	-0.16	0.08	-0.04	0.11	0.17*	-0.04	0.13	-0.19*	0.07	-0.06	0.05	-0.04	-0.04	-0.25**	-0.13	0.09	-0.03
ConfAbs3	0.18*	-0.16	0.11	0.01	0.2*	0.35**	-0.07	0.2*	-0.28**	0.14	-0.11	0.06	0.04	0.01	-0.21*	-0.05	0.18*	0.06
ConfAbs4	0.21*	-0.15	0.2*	0.03	0.13	0.34**	-0.06	0.21*	-0.28**	0.07	-0.12	0.05	-0.09	-0.09	-0.17	-0.12	0.09	-0.02
ConfAbsOverallPost	0.17	-0.21*	0.16	-0.04	0.13	0.28**	0.06	0.17*	-0.28**	0.08	-0.07	0.06	-0.03	-0.05	-0.17	-0.14	0.17	0.09

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 10 – *cont'd.*

	extroGoldberg	neurotGoldberg	intelGoldberg	agreeGoldberg	conGoldberg	pvIPC7	nvIPC7	pemIPC7	nemIPC7	conIPC7	agreeIPC7	cnvIPC7	ReaRIASEC	InvRIASEC	ArtRIASEC	SocRIASEC	EntRIASEC	ConRIASEC
ConfRelOverallPre	0.3**	-0.19*	0.27**	-0.09	0.16	0.35**	0.03	0.16	-0.22*	0.06	-0.23*	-0.09	-0.03	0.03	-0.16	-0.19*	0.03	0.08
ConfRel1	0.26**	-0.18*	0.19*	-0.03	0.15	0.33**	-0.01	0.17	-0.19*	0.06	-0.22*	0.09	0	0.01	-0.1	-0.1	0.01	-0.02
ConfRel2	0.25**	-0.17	0.07	-0.06	0.19*	0.3**	0	0.18*	-0.16	0.1	-0.16	0.14	-0.02	0.03	-0.21*	-0.16	0.08	0.01
ConfRel3	0.24**	-0.22*	0.07	-0.03	0.24**	0.32**	-0.04	0.21*	-0.31**	0.18*	-0.15	0.1	0.06	-0.02	-0.18*	-0.05	0.09	0.05
ConfRel4	0.29**	-0.22*	0.2*	0.01	0.17	0.37**	-0.04	0.27**	-0.33**	0.13	-0.16	0.1	-0.07	-0.07	-0.21*	-0.12	0.11	0.01
ConfRelOverallPost	0.18*	-0.17	0.12	-0.19*	0.13	0.26**	0.12	0.09	-0.29**	0.08	-0.2*	0.07	0.13	0.1	-0.15	-0.16	0.18*	0.2*
ACT_COMP_SCR	-0.09	-0.15	0.25**	-0.23*	-0.06	-0.09	-0.01	-0.1	-0.12	-0.17	-0.01	-0.21*	0.18	0.29**	0.22*	-0.16	-0.19*	0.08
ACT_ENGL_SCR	-0.11	-0.15	0.26**	-0.17	-0.09	-0.11	-0.01	-0.09	-0.11	-0.15	0.1	-0.29**	0.08	0.22*	0.28**	-0.05	-0.16	0.05
ACT_ENGWR_SCR	-0.12	-0.19	0.26**	-0.16	-0.09	-0.09	-0.05	-0.07	-0.13	-0.17	0.11	-0.27**	0	0.22*	0.26**	-0.08	-0.2*	-0.1
ACT_MATH_SCR	-0.08	-0.12	0.2*	-0.2*	-0.03	-0.02	0.02	-0.12	-0.08	-0.13	-0.12	-0.09	0.27**	0.35**	0.16	-0.2*	-0.13	0.09
ACT_READ_SCR	-0.03	-0.16	0.27**	-0.21*	0	-0.05	-0.02	-0.02	-0.1	-0.11	0.02	-0.18	0.09	0.17	0.17	-0.14	-0.14	0.04
ACT_SCIRE_SCR	-0.06	-0.09	0.1	-0.29**	-0.13	-0.03	0.02	-0.08	-0.11	-0.27**	-0.15	-0.21*	0.24*	0.33**	0.09	-0.2*	-0.26**	0.14
CUM_GPA	0.07	-0.03	0.14	0.04	0.11	0.02	-0.21*	0.08	0.1	0.09	-0.06	-0.04	-0.12	0.06	0.03	-0.05	-0.06	-0.08
TOT_ACAD_HOURS	-0.01	-0.07	0.12	-0.04	0.05	-0.04	-0.03	-0.04	-0.07	-0.01	-0.02	-0.13	0.03	-0.02	0.04	-0.04	-0.19*	-0.01
HS_RANK_PCT	0.01	-0.05	0.18	-0.12	-0.02	-0.09	0.18	-0.05	-0.01	-0.13	-0.12	-0.07	0.07	0.18	0.12	-0.24*	-0.23*	0.11
Gender	0.03	-0.2*	-0.06	-0.16	-0.01	0.23**	0.1	-0.12	-0.19*	-0.1	-0.25**	-0.01	0.37**	0.15	-0.24**	-0.31**	-0.13	0.08

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 10 – *cont'd.*

	extroGoldberg	neurotGoldberg	intelGoldberg	agreeGoldberg	conGoldberg	pvIPC7	nvIPC7	pemIPC7	nemIPC7	conIPC7	agreeIPC7	cnvIPC7	ReaRIASEC	InvRIASEC	ArtRIASEC	SocRIASEC	EntRIASEC	ConRIASEC
extroGoldberg	1**	-0.3**	0.14	0.18*	0.17*	0.45**	-0.2*	0.87**	-0.32**	-0.06	-0.14	0.06	-0.18*	-0.03	-0.01	0.08	0.12	-0.11
neurotGoldberg	-0.3**	1**	-0.17*	-0.3**	-0.19*	-0.28**	0.23**	-0.2*	0.76**	0.05	-0.21*	-0.01	-0.02	-0.08	0.01	0.16	0.08	0.14
intelGoldberg	0.14	-0.17*	1**	0.35**	0.23**	0.28**	-0.2*	0.13	-0.19*	-0.05	0.12	-0.33**	-0.06	0.08	0.34**	0	-0.15	-0.18*
agreeGoldberg	0.18*	-0.3**	0.35**	1**	0.35**	0.21*	-0.57**	0.28**	-0.27**	0.15	0.49**	0.08	-0.25**	-0.06	0.08	0.1	-0.09	-0.22**
conGoldberg	0.17*	-0.19*	0.23**	0.35**	1**	0.31**	-0.29**	0.13	-0.22**	0.7**	-0.04	0.35**	-0.18*	0.01	-0.06	0.06	0.2*	-0.03
pvIPC7	0.45**	-0.28**	0.28**	0.21*	0.31**	1**	-0.21*	0.37**	-0.34**	0.19*	-0.12	0.09	-0.15	0.02	-0.06	0.02	0.18*	-0.06
nvIPC7	-0.2*	0.23**	-0.2*	-0.57**	-0.29**	-0.21*	1**	-0.25**	0.25**	-0.04	-0.35**	-0.16	0.24**	-0.04	-0.04	-0.14	0.05	0.27**
pemIPC7	0.87**	-0.2*	0.13	0.28**	0.13	0.37**	-0.25**	1**	-0.25**	-0.08	0.05	0.02	-0.33**	-0.14	-0.05	0.14	0.11	-0.16
nemIPC7	-0.32**	0.76**	-0.19*	-0.27**	-0.22**	-0.34**	0.25**	-0.25**	1**	0.01	-0.2*	-0.01	-0.06	-0.12	-0.01	-0.01	-0.03	0.04
conIPC7	-0.06	0.05	-0.05	0.15	0.7**	0.19*	-0.04	-0.08	0.01	1**	-0.01	0.49**	-0.11	-0.05	-0.16	0.08	0.32**	0.16
agreeIPC7	-0.14	-0.21*	0.12	0.49**	-0.04	-0.12	-0.35**	0.05	-0.2*	-0.01	1**	-0.09	-0.12	0	0.12	0.08	-0.07	-0.1
cnvIPC7	0.06	-0.01	-0.33**	0.08	0.35**	0.09	-0.16	0.02	-0.01	0.49**	-0.09	1**	-0.13	-0.1	-0.29**	-0.02	0.15	-0.06
ReaRIASEC	-0.18*	-0.02	-0.06	-0.25**	-0.18*	-0.15	0.24**	-0.33**	-0.06	-0.11	-0.12	-0.13	1**	0.44**	0.26**	0.02	0.12	0.45**
InvRIASEC	-0.03	-0.08	0.08	-0.06	0.01	0.02	-0.04	-0.14	-0.12	-0.05	0	-0.1	0.44**	1**	0.32**	0.18*	-0.03	0.24**
ArtRIASEC	-0.01	0.01	0.34**	0.08	-0.06	-0.06	-0.04	-0.05	-0.01	-0.16	0.12	-0.29**	0.26**	0.32**	1**	0.39**	0.1	0.1
SocRIASEC	0.08	0.16	0	0.1	0.06	0.02	-0.14	0.14	-0.01	0.08	0.08	-0.02	0.02	0.18*	0.39**	1**	0.23**	0.11
EntRIASEC	0.12	0.08	-0.15	-0.09	0.2*	0.18*	0.05	0.11	-0.03	0.32**	-0.07	0.15	0.12	-0.03	0.1	0.23**	1**	0.47**
ConRIASEC	-0.11	0.14	-0.18*	-0.22**	-0.03	-0.06	0.27**	-0.16	0.04	0.16	-0.1	-0.06	0.45**	0.24**	0.1	0.11	0.47**	1**

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 10 – *cont'd.*

	extroGoldberg	neurotGoldberg	intelGoldberg	agreeGoldberg	conGoldberg	pvIPC7	nvIPC7	pemIPC7	nemIPC7	conIPC7	agreeIPC7	cnvIPC7	ReaRIASEC	InvRIASEC	ArtRIASEC	SocRIASEC	EntRIASEC	ConRIASEC
numpredout	0.13	0.03	0.03	-0.01	0.1	0.16	-0.03	0.14	-0.03	-0.01	0.03	0.03	-0.04	0.01	-0.18*	0.01	0.04	-0.12
numratees	0.17*	-0.1	0.09	-0.03	0.12	0.14	-0.04	0.14	-0.08	-0.04	-0.01	-0.08	0	-0.03	-0.01	0.03	0	-0.1
lengthtimepred	0.13	0.02	0.01	-0.11	0.06	0	0.08	0.13	0	0.01	-0.11	0.01	0	0	-0.03	0.06	-0.01	-0.08
numpredout formal	0.19*	-0.21*	0	0.07	0.08	0.15	-0.1	0.06	-0.19*	0.02	-0.05	0	0.1	0.05	-0.06	-0.04	0.04	-0.07
numpredout training	0.2*	0.03	0.13	-0.06	-0.02	0.1	-0.08	0.2*	-0.05	-0.03	-0.05	-0.09	-0.04	-0.07	-0.14	0.01	0.11	-0.01
numpredout trainingformal	0.12	0.01	0.19*	0.04	-0.06	0.08	-0.1	0.09	-0.09	-0.07	0.01	-0.12	0.04	-0.06	-0.12	0	0.08	0.02
lengthtraining	0.11	0.04	-0.13	-0.07	-0.08	-0.02	0.03	0.04	0.02	-0.05	-0.17*	0.09	0.1	-0.02	-0.06	0.12	0.04	-0.04
amtstats	-0.05	0.07	-0.03	-0.07	0.03	0.05	0.06	-0.04	-0.03	0.01	-0.05	-0.16	-0.02	-0.06	0.05	0.06	0.13	0.16
amtdecision mkg	0.16	-0.04	-0.06	0	-0.01	0.1	0	0.12	-0.03	0.05	-0.12	-0.01	-0.07	-0.15	-0.15	-0.08	0.15	0.1
TrPerExprAware	0.2*	-0.04	0.19*	0.17	0.14	0.05	-0.12	0.18*	-0.03	-0.04	0.01	-0.14	-0.1	0.04	0.09	0.05	-0.11	-0.12
TrPerSaidUsed	0.13	-0.01	0.19*	0.12	0.07	0.02	-0.12	0.12	-0.01	-0.08	0.01	-0.18*	-0.02	0.09	0.14	0.07	-0.21*	-0.16
TrPerUsedCorrectly	0.14	-0.15	0.1	0.16	0.15	0.03	-0.12	0.13	-0.15	0.01	0.09	-0.09	-0.17*	0.11	-0.02	-0.01	-0.18*	-0.15
TrPerSaidUsed3rdVar	0.02	-0.06	0.05	0.15	0.19*	0.07	-0.06	0.05	-0.1	0.05	0.02	-0.02	0.03	0.14	-0.04	-0.07	0.04	-0.01

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 10 – *cont'd.*

	numpredout	numratees	lengthtimepred	numpredoutformal	numpredouttraining	numpredouttrainingformal	lengthtraining	amtstats	amtdecisionmkg	TrPerExprAware	TrPerSaidUsed	TrPerUsedCorrectly	TrPerSaidUsed3rdVar
Fisherra	-0.01	0.12	-0.02	0.03	-0.06	-0.11	-0.16	0	-0.15	0.13	0.15	0.18*	0.04
Fisherra1	-0.06	0.06	-0.11	0.11	0.04	-0.09	-0.07	0.05	0.06	0.16	0.11	0.08	0.12
Fisherra2	0.08	0.09	0.01	0.08	0.07	-0.07	-0.19*	0.04	-0.2*	0.09	0.09	0.14	0.06
Fisherra3	0.05	0.14	0.09	0.04	-0.01	-0.04	-0.1	-0.09	-0.16	0	0.06	0.13	0.06
Fisherra4	-0.01	0.09	-0.02	-0.09	-0.11	-0.08	-0.11	-0.02	-0.13	0.08	0.13	0.12	-0.04
SS	0.1	0.15	-0.07	0.14	0.01	-0.14	-0.06	0.02	-0.03	0	0.05	0.06	0.08
SS1	0.13	0.19*	0.02	0.21*	0.04	-0.01	0.04	0.07	0.14	0.01	-0.05	0.04	0.06
SS2	0.05	0.05	-0.1	0.12	0.01	-0.16	-0.06	0.06	-0.05	-0.03	0.02	-0.04	0.06
SS3	0.11	0.12	0.01	0.08	0.08	-0.06	-0.09	-0.02	-0.05	0	0.05	0.1	0.07
SS4	0.02	0.08	-0.11	0	-0.06	-0.17*	-0.09	-0.06	-0.15	0.02	0.11	0.1	0.06
FisherRslin	-0.01	-0.06	-0.03	-0.14	-0.03	0.05	-0.12	0.04	-0.08	-0.15	-0.22**	-0.04	0.01
FisherRslin1	-0.09	-0.04	-0.11	-0.14	-0.14	-0.07	-0.14	-0.02	-0.07	-0.21*	-0.2*	-0.04	-0.07
FisherRslin2	-0.07	-0.01	-0.04	-0.13	-0.03	-0.02	-0.13	-0.09	-0.1	-0.17*	-0.18*	-0.04	-0.03
FisherRslin3	0.03	0	-0.03	-0.11	0.02	0.04	-0.07	-0.05	-0.07	-0.1	-0.14	-0.1	-0.03
FisherRslin4	-0.01	-0.09	-0.05	-0.1	-0.03	0.05	-0.07	0	-0.05	-0.12	-0.19*	-0.13	-0.02
FisherG	0.04	0.12	0.02	-0.16	-0.17*	-0.22**	-0.12	0.09	-0.03	-0.19*	-0.2*	-0.03	-0.02
FisherG1	0.03	0.14	0	-0.03	-0.21*	-0.22**	-0.09	0.01	0.03	-0.16	-0.16	0.04	0.03
FisherG2	0.04	0.08	-0.03	0.01	0	-0.09	-0.11	-0.07	-0.17*	0.02	0.02	0.09	-0.06
FisherG3	0.15	0.19*	0.1	-0.07	0.05	-0.13	-0.05	-0.06	0.01	-0.1	-0.11	-0.07	-0.02
FisherG4	0.09	-0.06	0.05	-0.13	-0.09	-0.14	-0.08	0.04	-0.09	-0.14	-0.11	0.02	-0.08

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 10 – *cont'd.*

	numpredout	numratees	lengthtimepred	numpredoutformal	numpredouttraining	numpredouttrainingformal	lengthtraining	amtstats	amtdecisionmkg	TrPerExprAware	TrPerSaidUsed	TrPerUsedCorrectly	TrPerSaidUsed3rdVar
FisherCstat	-0.09	0.06	-0.09	0.05	0.01	0.03	-0.1	0.03	-0.13	0.23**	0.24**	0.16	0.09
FisherCstat1	-0.1	0.04	-0.12	0.08	0.17*	0.1	-0.02	0.07	0	0.23**	0.18*	0.11	0.12
FisherCstat2	-0.02	0.06	-0.06	0.12	0.05	-0.03	-0.12	0.05	-0.19*	0.05	0.07	0.06	0.12
FisherCstat3	0	0.08	0.08	0.04	-0.06	-0.06	-0.1	-0.07	-0.11	0.03	0.09	0.15	0.1
FisherCstat4	-0.06	0.07	-0.1	-0.06	-0.08	-0.04	-0.07	0	-0.11	0.13	0.16	0.09	0.02
Fisherrz	-0.09	0.06	-0.09	0.05	0.01	0.03	-0.1	0.03	-0.13	0.23**	0.24**	0.16	0.09
Fisherrz1	-0.1	0.04	-0.12	0.08	0.17*	0.1	-0.02	0.07	0	0.23**	0.18*	0.11	0.12
Fisherrz2	-0.02	0.05	-0.06	0.12	0.04	-0.03	-0.12	0.05	-0.19*	0.05	0.07	0.06	0.13
Fisherrz3	0	0.08	0.08	0.04	-0.06	-0.06	-0.1	-0.07	-0.11	0.03	0.09	0.15	0.1
Fisherrz4	-0.06	0.07	-0.1	-0.06	-0.08	-0.04	-0.07	0	-0.11	0.13	0.16	0.09	0.02
Cwtx1x2_s	0.07	0.06	0.04	-0.05	-0.01	-0.05	-0.11	0.06	-0.04	-0.07	-0.11	0.13	-0.02
Cwtx1x2_1_s	-0.04	0.08	-0.09	-0.01	0.02	-0.02	-0.13	0.03	0.01	0.03	0.11	0.13	0.09
Cwtx1x2_2_s	0.07	0.08	0	-0.02	0.05	0.06	-0.19*	0.06	-0.02	0.04	-0.05	0.17	0.04
Cwtx1x2_3_s	-0.05	0.04	0.01	-0.14	0.01	0.02	-0.18*	0.03	-0.01	-0.07	-0.06	0.19*	-0.06
Cwtx1x2_4_s	0.02	0	0	-0.08	0.09	0.04	-0.05	0.02	0.04	0.18*	0.2*	0.35**	-0.13
ConfAbsOverallPre	0.21*	0.13	0.13	0.07	0.19*	0.2*	-0.08	0.1	-0.02	0	-0.04	-0.07	0.14
ConfAbs1	0.04	-0.03	0.13	-0.01	0.08	0.09	-0.08	0.02	0.08	0.05	0.06	-0.08	0.06
ConfAbs2	0.12	-0.02	0.1	-0.03	0.08	0.06	-0.05	-0.06	0.02	-0.04	-0.05	-0.05	0.02
ConfAbs3	0.1	-0.03	0.13	0.01	0.14	0.05	-0.01	0.03	0.07	0.01	-0.03	-0.05	0.15
ConfAbs4	0.04	-0.05	0.1	0.03	0.08	0.06	0.02	-0.01	0.16	0.02	0.01	-0.05	0.18
ConfAbsOverallPost	0.07	0	0.12	-0.08	0.07	0.1	-0.07	0.04	-0.03	-0.04	-0.1	-0.13	0.18

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 10 – *cont'd.*

	numpredout	numratees	lengthtimepred	numpredoutformal	numpredouttraining	numpredouttrainingformal	lengthtraining	amtstats	amtdecisionmkg	TrPerExprAware	TrPerSaidUsed	TrPerUsedCorrectly	TrPerSaidUsed3rdVar
ConfRelOverallPre	0.24*	0.18	0.17	0.14	0.21*	0.13	-0.04	0.12	0.12	0.08	0.04	-0.05	0.11
ConfRel1	0.04	0.06	0.05	0.07	0.05	0.06	-0.02	0.1	0.07	0.04	0.05	-0.12	0.07
ConfRel2	0.12	0.05	0.04	0.04	0.03	0.07	-0.03	0.02	0.07	-0.04	-0.07	-0.14	0.13
ConfRel3	0.05	0.01	0.04	0.08	0.13	0.06	0.03	0.06	0.09	0.03	0.02	-0.04	0.1
ConfRel4	-0.03	-0.04	0	0.01	0.04	0.01	-0.06	0.01	0.11	-0.01	-0.04	-0.12	0.13
ConfRelOverallPost	0.07	0.02	0.07	0.02	0.1	0.09	0.01	0.11	0.03	-0.04	-0.09	-0.15	0.19*
ACT_COMP_SCR	-0.07	-0.07	-0.04	-0.16	0.08	-0.05	-0.1	-0.06	-0.19*	0.02	0.17	0.23*	-0.01
ACT_ENGL_SCR	-0.09	0	-0.06	-0.18	0.11	-0.02	-0.14	0.03	-0.17	0.04	0.13	0.21*	0.02
ACT_ENGWR_SCR	0	0.04	0	-0.06	0.21*	0.06	-0.07	0.1	-0.07	0.09	0.17	0.27**	0.03
ACT_MATH_SCR	-0.05	-0.1	-0.03	-0.14	0.01	-0.09	-0.05	-0.14	-0.21*	-0.03	0.08	0.16	0.03
ACT_READ_SCR	-0.07	-0.1	-0.02	-0.13	0.09	-0.08	-0.1	-0.01	-0.13	0.1	0.25**	0.23*	-0.06
ACT_SCIRE_SCR	0.04	0.06	0.04	-0.08	0.13	0.05	0	-0.04	-0.19	0.03	0.18	0.15	0
CUM_GPA	-0.07	0.04	0.01	-0.01	-0.06	-0.07	-0.06	0	0.01	0.04	0.13	0.22*	-0.09
TOT_ACAD_HOURS	0.16	0.25**	0.04	0.31**	0.21*	0.16	0.08	0.4**	0.11	0.11	0.13	0.05	0.07
HS_RANK_PCT	0.02	0.09	0.1	-0.08	0.02	-0.09	-0.01	0	-0.19	0	0.06	0.06	-0.14
Gender	0.01	0.02	0.01	0.19*	-0.05	0.04	0.11	-0.05	0.13	-0.02	0.02	0	0.09

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 10 – *cont'd.*

	numpredout	numratees	lengthtimepred	numpredoutformal	numpredouttraining	numpredouttrainingformal	lengthtraining	amtstats	amtdecisionmkg	TrPerExprAware	TrPerSaidUsed	TrPerUsedCorrectly	TrPerSaidUsed3rdVar
extroGoldberg	0.13	0.17*	0.13	0.19*	0.2*	0.12	0.11	-0.05	0.16	0.2*	0.13	0.14	0.02
neurotGoldberg	0.03	-0.1	0.02	-0.21*	0.03	0.01	0.04	0.07	-0.04	-0.04	-0.01	-0.15	-0.06
intelGoldberg	0.03	0.09	0.01	0	0.13	0.19*	-0.13	-0.03	-0.06	0.19*	0.19*	0.1	0.05
agreeGoldberg	-0.01	-0.03	-0.11	0.07	-0.06	0.04	-0.07	-0.07	0	0.17	0.12	0.16	0.15
conGoldberg	0.1	0.12	0.06	0.08	-0.02	-0.06	-0.08	0.03	-0.01	0.14	0.07	0.15	0.19*
pvIPC7	0.16	0.14	0	0.15	0.1	0.08	-0.02	0.05	0.1	0.05	0.02	0.03	0.07
nvIPC7	-0.03	-0.04	0.08	-0.1	-0.08	-0.1	0.03	0.06	0	-0.12	-0.12	-0.12	-0.06
pemIPC7	0.14	0.14	0.13	0.06	0.2*	0.09	0.04	-0.04	0.12	0.18*	0.12	0.13	0.05
nemIPC7	-0.03	-0.08	0	-0.19*	-0.05	-0.09	0.02	-0.03	-0.03	-0.03	-0.01	-0.15	-0.1
conIPC7	-0.01	-0.04	0.01	0.02	-0.03	-0.07	-0.05	0.01	0.05	-0.04	-0.08	0.01	0.05
agreeIPC7	0.03	-0.01	-0.11	-0.05	-0.05	0.01	-0.17*	-0.05	-0.12	0.01	0.01	0.09	0.02
cnvIPC7	0.03	-0.08	0.01	0	-0.09	-0.12	0.09	-0.16	-0.01	-0.14	-0.18*	-0.09	-0.02
ReaRIASEC	-0.04	0	0	0.1	-0.04	0.04	0.1	-0.02	-0.07	-0.1	-0.02	-0.17*	0.03
InvRIASEC	0.01	-0.03	0	0.05	-0.07	-0.06	-0.02	-0.06	-0.15	0.04	0.09	0.11	0.14
ArtRIASEC	-0.18*	-0.01	-0.03	-0.06	-0.14	-0.12	-0.06	0.05	-0.15	0.09	0.14	-0.02	-0.04
SocRIASEC	0.01	0.03	0.06	-0.04	0.01	0	0.12	0.06	-0.08	0.05	0.07	-0.01	-0.07
EntRIASEC	0.04	0	-0.01	0.04	0.11	0.08	0.04	0.13	0.15	-0.11	-0.21*	-0.18*	0.04
ConRIASEC	-0.12	-0.1	-0.08	-0.07	-0.01	0.02	-0.04	0.16	0.1	-0.12	-0.16	-0.15	-0.01

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 10 – *cont'd.*

	numpredout	numratees	lengthtimepred	numpredoutformal	numpredouttraining	numpredouttrainingformal	lengthtraining	amtstats	amtdecisionmkg	TrPerExprAware	TrPerSaidUsed	TrPerUsedCorrectly	TrPerSaidUsed3rdVar
numpredout	1**	0.58**	0.57**	0.36**	0.5**	0.4**	0.23**	0.18*	0.08	0.12	0.01	0.02	0.12
numratees	0.58**	1**	0.39**	0.38**	0.26**	0.27**	0.04	0.2*	-0.03	0.11	0.07	0.06	0.21*
lengthtimepred	0.57**	0.39**	1**	0.18*	0.28**	0.17*	0.31**	0.06	0.11	0.01	0.01	0.01	-0.05
numpredoutformal	0.36**	0.38**	0.18*	1**	0.29**	0.38**	0.38**	0.13	0.27**	0.02	-0.02	0.11	0.03
numpredouttraining	0.5**	0.26**	0.28**	0.29**	1**	0.63**	0.28**	0.1	0.12	0.19*	0.13	0.09	-0.01
numpredouttrainingformal	0.4**	0.27**	0.17*	0.38**	0.63**	1**	0.14	0.17*	0.09	0.15	0.04	0.07	-0.06
lengthtraining	0.23**	0.04	0.31**	0.38**	0.28**	0.14	1**	-0.02	0.31**	-0.04	-0.02	-0.04	-0.07
Amtstats	0.18*	0.2*	0.06	0.13	0.1	0.17*	-0.02	1**	0.25**	-0.02	0.01	0.06	0.16
amtdecisionmkg	0.08	-0.03	0.11	0.27**	0.12	0.09	0.31**	0.25**	1**	0	-0.01	0.03	-0.09
TrPerExprAware	0.12	0.11	0.01	0.02	0.19*	0.15	-0.04	-0.02	0	1**	0.88**	0.41**	0.15
TrPerSaidUsed	0.01	0.07	0.01	-0.02	0.13	0.04	-0.02	0.01	-0.01	0.88**	1**	0.39**	0.16
TrPerUsedCorrectly	0.02	0.06	0.01	0.11	0.09	0.07	-0.04	0.06	0.03	0.41**	0.39**	1**	0.1
TrPerSaidUsed3rdVar	0.12	0.21*	-0.05	0.03	-0.01	-0.06	-0.07	0.16	-0.09	0.15	0.16	0.1	1**

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 11. *Incremental Prediction of Change Over Time in Judgment Validity (r_a): Individual Differences Variables (Simpler Model = Random Intercepts, Random Slopes, No Time x Feedback interaction) (Fall 2009)*

		Fix Eff Ind Diff	Std Err Fix Eff	t-val Fix Eff	t-val 95% (|t| >2)	df simpler model	df more complex model	AIC simpler model	AIC more complex model	BIC simpler model	BIC more complex model	logLik simpler model	logLik more complex mod	Chisq	Chi df	pval (Chisq)	sig
1	ACT_COMP_SCR	0.0069	0.0021	3.26	!	6	7	-555.15	-434.84	-529.09	-406.17	283.57	224.42	0.00	1	1.00E+00	
2	ACT_ENGL_SCR	0.0072	0.0017	4.19	!	6	7	-555.15	-441.35	-529.09	-412.68	283.57	227.67	0.00	1	1.00E+00	
3	ACT_ENGWR_SCR	0.0084	0.0020	4.22	!	6	7	-555.15	-400.70	-529.09	-372.56	283.57	207.35	0.00	1	1.00E+00	
4	ACT_MATH_SCR	0.0036	0.0020	1.78		6	7	-555.15	-428.34	-529.09	-399.67	283.57	221.17	0.00	1	1.00E+00	
5	ACT_READ_SCR	0.0045	0.0016	2.73	!	6	7	-555.15	-432.40	-529.09	-403.73	283.57	223.20	0.00	1	1.00E+00	
6	ACT_SCIRE_SCR	0.0049	0.0022	2.28	!	6	7	-555.15	-430.08	-529.09	-401.41	283.57	222.04	0.00	1	1.00E+00	
7	CUM_GPA	0.0340	0.0163	2.08	!	6	7	-555.15	-514.11	-529.09	-484.12	283.57	264.06	0.00	1	1.00E+00	
8	TOT_ACAD_HOURS	0.0004	0.0002	1.59		6	7	-555.15	-512.40	-529.09	-482.41	283.57	263.20	0.00	1	1.00E+00	
9	HS_RANK_PCT	0.0010	0.0004	2.31	!	6	7	-555.15	-362.11	-529.09	-334.38	283.57	188.05	0.00	1	1.00E+00	
10	Gender	-0.0067	0.0175	-0.38		6	7	-555.15	-548.67	-529.09	-518.32	283.57	281.33	0.00	1	1.00E+00	
11	extroGoldberg	-0.0009	0.0004	-2.65	!	6	7	-555.15	-560.08	-529.09	-529.69	283.57	287.04	6.94	1	8.44E-03	**
12	neurotGoldberg	-0.0005	0.0004	-1.17		6	7	-555.15	-554.51	-529.09	-524.12	283.57	284.26	1.37	1	2.43E-01	
13	intelGoldberg	0.0007	0.0005	1.30		6	7	-555.15	-554.83	-529.09	-524.44	283.57	284.42	1.68	1	1.94E-01	
14	agreeGoldberg	0.0002	0.0005	0.44		6	7	-555.15	-553.33	-529.09	-522.94	283.57	283.67	0.19	1	6.66E-01	
15	conGoldberg	-0.0002	0.0004	-0.38		6	7	-555.15	-553.29	-529.09	-522.89	283.57	283.64	0.14	1	7.10E-01	

Notes. Fix Eff = fixed effect; Ind Diff = individual difference; Std Err = standard error; t-val = t-value; df = degrees of freedom; AIC = Akaike information criterion, BIC = Bayesian information criterion; logLik = log likelihood; chi or chisq = chi-square χ^2; pval = probability value; simpler model = regression model that best fits the data without the individual difference variable included; more complex model = simpler model with the individual difference included; ! = absolute value for the t-value is greater than 2 (±1.96 = z-score for 2-tailed z-test with $\alpha = 0.05$); ** = p-value < 0.01; * = p-value < 0.05.

Table 11 – *cont'd.*

		Fix Eff Ind Diff	Std Err Fix Eff	t-val Fix Eff	t-val 95% (|t| >2)	df simpler model	df more complex model	AIC simpler model	AIC more complex model	BIC simpler model	BIC more complex model	logLik simpler model	logLik more complex mod	Chisq	Chi df	pval (Chisq)	sig
16	pvIPC7	-0.0022	0.0016	-1.33		6	7	-555.15	-554.90	-529.09	-524.51	283.57	284.45	1.75	1	1.85E-01	
17	nvIPC7	-0.0006	0.0025	-0.22		6	7	-555.15	-553.19	-529.09	-522.79	283.57	283.59	0.04	1	8.37E-01	
18	pemIPC7	-0.0032	0.0015	-2.14	!	6	7	-555.15	-557.71	-529.09	-527.32	283.57	285.86	4.57	1	3.26E-02	*
19	nemIPC7	-0.0011	0.0018	-0.62		6	7	-555.15	-553.53	-529.09	-523.14	283.57	283.77	0.39	1	5.34E-01	
20	conIPC7	-0.0022	0.0017	-1.28		6	7	-555.15	-554.72	-529.09	-524.32	283.57	284.36	1.57	1	2.10E-01	
21	agreeIPC7	0.0034	0.0021	1.61		6	7	-555.15	-555.75	-529.09	-525.35	283.57	284.87	2.60	1	1.07E-01	
22	cnvIPC7	-0.0036	0.0018	-1.97		6	7	-555.15	-557.01	-529.09	-526.61	283.57	285.50	3.86	1	4.95E-02	*
23	ReaRIASEC	-0.0005	0.0009	-0.55		6	7	-555.15	-553.44	-529.09	-523.05	283.57	283.72	0.30	1	5.87E-01	
24	InvRIASEC	0.0006	0.0008	0.76		6	7	-555.15	-553.73	-529.09	-523.34	283.57	283.87	0.58	1	4.45E-01	
25	ArtRIASEC	0.0005	0.0009	0.57		6	7	-555.15	-553.47	-529.09	-523.07	283.57	283.73	0.32	1	5.71E-01	
26	SocRIASEC	-0.0015	0.0011	-1.29		6	7	-555.15	-554.81	-529.09	-524.41	283.57	284.40	1.66	1	1.97E-01	
27	EntRIASEC	-0.0019	0.0010	-1.96		6	7	-555.15	-556.93	-529.09	-526.53	283.57	285.46	3.78	1	5.19E-02	
28	ConRIASEC	0.0018	0.0009	2.06	!	6	7	-555.15	-557.36	-529.09	-526.97	283.57	285.68	4.21	1	4.01E-02	*
29	numpredout	0.0018	0.0061	0.30		6	7	-555.15	-553.23	-529.09	-522.84	283.57	283.62	0.09	1	7.70E-01	
30	numratees	0.0004	0.0003	1.65		6	7	-555.15	-555.88	-529.09	-525.49	283.57	284.94	2.73	1	9.82E-02	
31	lengthtimepred	0.0000	0.0002	-0.19		6	7	-555.15	-553.18	-529.09	-522.78	283.57	283.59	0.03	1	8.60E-01	
32	numpredoutformal	0.0172	0.0163	1.06		6	7	-555.15	-554.25	-529.09	-523.85	283.57	284.12	1.10	1	2.94E-01	
33	numpredouttraining	0.0042	0.0122	0.34		6	7	-555.15	-553.26	-529.09	-522.86	283.57	283.63	0.11	1	7.37E-01	
34	numpredouttrainingformal	-0.0225	0.0186	-1.21		6	7	-555.15	-554.62	-529.09	-524.22	283.57	284.31	1.47	1	2.25E-01	
35	lengthtraining	-0.0003	0.0002	-2.14	!	6	7	-555.15	-557.72	-529.09	-527.32	283.57	285.86	4.57	1	3.25E-02	*
36	amtstats	0.0020	0.0162	0.12		6	7	-555.15	-553.16	-529.09	-522.76	283.57	283.58	0.01	1	9.24E-01	
37	amtdecisionmkg	-0.0745	0.0412	-1.81		6	7	-555.15	-556.38	-529.09	-525.98	283.57	285.19	3.23	1	7.23E-02	

Notes. Fix Eff = fixed effect; Ind Diff = individual difference; Std Err = standard error; t-val = t-value; df = degrees of freedom; AIC = Akaike information criterion, BIC = Bayesian information criterion; logLik = log likelihood; chi or chisq = chi-square χ^2; pval = probability value; simpler model = regression model that best fits the data without the individual difference variable included; more complex model = simpler model with the individual difference included; ! = absolute value for the t-value is greater than 2 (±1.96 = z-score for 2-tailed z-test with $\alpha = 0.05$); ** = p-value < 0.01; * = p-value < 0.05.

Table 12. *Incremental Prediction of Change Over Time in Skill Score: Individual Differences Variables (Simpler Model = Random Intercepts, Random Slopes, Time x Feedback Interaction) (Fall 2009)*

		Fix Eff Ind Diff	Std Err Fix Eff	t-val Fix Eff	t-val 95% (|t|>2)	df simpler model	df more complex model	AIC simpler model	AIC more complex model	BIC simpler model	BIC more complex model	logLik simpler model	logLik more complex mod	Chisq	Chi df	pval (Chisq)	sig
1	ACT_COMP_SCR	0.047	0.019	2.52	!	10	11	1701.70	1327.43	1745.12	1372.48	-840.85	-652.71	376.27	1	8.07E-84	**
2	ACT_ENGL_SCR	0.037	0.016	2.35	!	10	11	1701.70	1327.99	1745.12	1373.04	-840.85	-652.99	375.71	1	1.07E-83	**
3	ACT_ENGWR_SCR	0.045	0.018	2.43	!	10	11	1701.70	1247.06	1745.12	1291.30	-840.85	-612.53	456.63	1	2.60E-101	**
4	ACT_MATH_SCR	0.040	0.018	2.25	!	10	11	1701.70	1328.64	1745.12	1373.70	-840.85	-653.32	375.05	1	1.49E-83	**
5	ACT_READ_SCR	0.029	0.014	2.02	!	10	11	1701.70	1329.50	1745.12	1374.55	-840.85	-653.75	374.20	1	2.28E-83	**
6	ACT_SCIRE_SCR	0.040	0.019	2.12	!	10	11	1701.70	1329.04	1745.12	1374.09	-840.85	-653.52	374.66	1	1.81E-83	**
7	CUM_GPA	0.139	0.138	1.01		10	11	1701.70	1620.11	1745.12	1667.23	-840.85	-799.05	83.59	1	6.09E-20	**
8	TOT_ACAD_HOURS	0.003	0.002	1.38		10	11	1701.70	1619.19	1745.12	1666.32	-840.85	-798.60	84.50	1	3.83E-20	**
9	HS_RANK_PCT	0.003	0.004	0.93		10	11	1701.70	1159.16	1745.12	1202.73	-840.85	-568.58	544.54	1	1.94E-120	**
10	Gender	0.065	0.143	0.45		10	11	1701.70	1687.02	1745.12	1734.70	-840.85	-832.51	16.68	1	4.43E-05	**
11	extroGoldberg	-0.004	0.003	-1.26		10	11	1701.70	1702.08	1745.12	1749.84	-840.85	-840.04	1.62	1	2.03E-01	
12	neurotGoldberg	-0.002	0.003	-0.65		10	11	1701.70	1703.28	1745.12	1751.04	-840.85	-840.64	0.42	1	5.17E-01	
13	intelGoldberg	-0.002	0.004	-0.36		10	11	1701.70	1703.58	1745.12	1751.35	-840.85	-840.79	0.12	1	7.34E-01	
14	agreeGoldberg	-0.005	0.004	-1.40		10	11	1701.70	1701.73	1745.12	1749.49	-840.85	-839.86	1.97	1	1.61E-01	
15	conGoldberg	-0.003	0.003	-0.82		10	11	1701.70	1703.02	1745.12	1750.78	-840.85	-840.51	0.68	1	4.09E-01	

Notes. Fix Eff = fixed effect; Ind Diff = individual difference; Std Err = standard error; t-val = t-value; df = degrees of freedom; AIC = Akaike information criterion, BIC = Bayesian information criterion; logLik = log likelihood; chi or chisq = chi-square χ^2; pval = probability value; simpler model = regression model that best fits the data without the individual difference variable included; more complex model = simpler model with the individual difference included; ! = absolute value for the t-value is greater than 2 (± 1.96 = z-score for 2-tailed z-test with $\alpha = 0.05$); ** = p-value < 0.01; * = p-value < 0.05.

Table 12 – *cont'd.*

	Fix Eff Ind Diff	Std Err Fix Eff	t-val Fix Eff	t-val 95% (|t|>2)	df simpl er model	df more compl ex model	AIC simpl er model	AIC more complex model	BIC simpler model	BIC more complex model	logLik simpler model	logLik more complex mod	Chisq	Chi df	pv al (C his q)	sig	
16	pvIPC7	-0.023	0.013	-1.70		10	11	1701.70	1700.78	1745.12	1748.55	-840.85	-839.39	2.91	1	8.78E-02	
17	nvIPC7	0.006	0.021	0.31		10	11	1701.70	1703.61	1745.12	1751.38	-840.85	-840.81	0.08	1	7.73E-01	
18	pemIPC7	-0.019	0.012	-1.53		10	11	1701.70	1701.32	1745.12	1749.08	-840.85	-839.66	2.38	1	1.23E-01	
19	nemIPC7	-0.015	0.015	-0.99		10	11	1701.70	1702.71	1745.12	1750.47	-840.85	-840.35	0.99	1	3.20E-01	
20	conIPC7	-0.026	0.014	-1.92		10	11	1701.70	1699.99	1745.12	1747.75	-840.85	-838.99	3.71	1	5.41E-02	
21	agreeIPC7	0.005	0.017	0.32		10	11	1701.70	1703.61	1745.12	1751.37	-840.85	-840.80	0.09	1	7.62E-01	
22	cnvIPC7	-0.038	0.015	-2.62	!	10	11	1701.70	1696.81	1745.12	1744.57	-840.85	-837.40	6.89	1	8.66E-03	**
23	ReaRIASEC	0.018	0.007	2.41	!	10	11	1701.70	1697.87	1745.12	1745.63	-840.85	-837.94	5.83	1	1.58E-02	*
24	InvRIASEC	0.013	0.006	2.07	!	10	11	1701.70	1699.41	1745.12	1747.17	-840.85	-838.70	4.29	1	3.83E-02	*
25	ArtRIASEC	0.008	0.007	1.06		10	11	1701.70	1702.57	1745.12	1750.33	-840.85	-840.28	1.13	1	2.88E-01	
26	SocRIASEC	0.009	0.010	0.97		10	11	1701.70	1702.76	1745.12	1750.53	-840.85	-840.38	0.94	1	3.33E-01	
27	EntRIASEC	-0.002	0.008	-0.21		10	11	1701.70	1703.66	1745.12	1751.42	-840.85	-840.83	0.04	1	8.47E-01	
28	ConRIASEC	-0.001	0.007	-0.08		10	11	1701.70	1703.70	1745.12	1751.47	-840.85	-840.85	0.00	1	1.00E+00	

Notes. Fix Eff = fixed effect; Ind Diff = individual difference; Std Err = standard error; t-val = t-value; df = degrees of freedom; AIC = Akaike information criterion, BIC = Bayesian information criterion; logLik = log likelihood; chi or chisq = chi-square χ^2; pval = probability value; simpler model = regression model that best fits the data without the individual difference variable included; more complex model = simpler model with the individual difference included; ! = absolute value for the t-value is greater than 2 (±1.96 = z-score for 2-tailed z-test with $\alpha = 0.05$); ** = p-value < 0.01; * = p-value < 0.05.

Table 12 – *cont'd.*

		Fix Eff Ind Diff	Std Err Fix Eff	t-val Fix Eff	t-val 95% (|t|>2)	df simpler model	df more complex model	AIC simpler model	AIC more complex model	BIC simpler model	BIC more complex model	logLik simpler model	logLik more complex mod	Chisq	Chi df	pval (Chisq)	sig
29	numpredout	0.065	0.050	1.30		10	11	1701.70	1702.01	1745.12	1749.77	-840.85	-840.00	1.69	1	1.93E-01	
30	numratees	0.004	0.002	1.89		10	11	1701.70	1700.10	1745.12	1747.87	-840.85	-839.05	3.59	1	5.80E-02	
31	lengthtimepred	-0.001	0.001	-0.58		10	11	1701.70	1703.36	1745.12	1751.13	-840.85	-840.68	0.33	1	5.63E-01	
32	numpredoutformal	0.268	0.132	2.03	!	10	11	1701.70	1699.63	1745.12	1747.40	-840.85	-838.82	4.07	1	4.38E-02	*
33	numpredouttraining	0.039	0.101	0.39		10	11	1701.70	1703.56	1745.12	1751.32	-840.85	-840.78	0.14	1	7.08E-01	
34	numpredouttrainingformal	-0.232	0.155	-1.50		10	11	1701.70	1701.43	1745.12	1749.19	-840.85	-839.71	2.27	1	1.32E-01	
35	lengthtraining	-0.001	0.001	-0.58		10	11	1701.70	1703.37	1745.12	1751.13	-840.85	-840.68	0.33	1	5.65E-01	
36	amtstats	0.071	0.133	0.53		10	11	1701.70	1703.43	1745.12	1751.19	-840.85	-840.71	0.27	1	6.03E-01	
37	amtdecisionmkg	-0.030	0.342	-0.09		10	11	1701.70	1703.70	1745.12	1751.46	-840.85	-840.85	0.00	1	1.00E+00	

Notes. Fix Eff = fixed effect; Ind Diff = individual difference; Std Err = standard error; t-val = t-value; df = degrees of freedom; AIC = Akaike information criterion, BIC = Bayesian information criterion; logLik = log likelihood; chi or chisq = chi-square χ^2; pval = probability value; simpler model = regression model that best fits the data without the individual difference variable included; more complex model = simpler model with the individual difference included; ! = absolute value for the t-value is greater than 2 (± 1.96 = z-score for 2-tailed z-test with $\alpha = 0.05$); ** = p-value < 0.01; * = p-value < 0.05.

Table 13. *Insight: Interrater Agreement (Fall 2009)*

Block	Dimension Rated	Disagreement (Absolute Frequencies)					Agreement (Relative Frequencies)			
		Overall	C1	C2	C3		Overall	C1	C2	C3
1	Expressed awareness of disordinal interaction?	8	4	2	2		94%	91%	95%	96%
	Said s/he used disordinal interaction?	10	3	3	4		93%	93%	93%	92%
	Used disordinal interaction properly?	19	7	5	7		86%	84%	88%	86%
2	Expressed awareness of disordinal interaction?	6	1	4	1		96%	98%	90%	98%
	Said s/he used disordinal interaction?	5	1	4	0		96%	98%	90%	100%
	Used disordinal interaction properly?	11	4	2	5		92%	91%	95%	90%
3	Expressed awareness of disordinal interaction?	5	1	1	3		96%	98%	98%	94%
	Said s/he used disordinal interaction?	4	1	1	2		97%	98%	98%	96%
	Used disordinal interaction properly?	10	2	4	4		93%	95%	90%	92%
4	Expressed awareness of disordinal interaction?	9	1	2	6		93%	98%	95%	88%
	Said s/he used disordinal interaction?	5	3	0	2		96%	93%	100%	96%
	Used disordinal interaction properly?	11	1	6	4		92%	98%	86%	92%
Overall (after all predictions were made)	With which block did s/he think his/her strategy changed (no change = 0)?	4	1	2	1		97%	98%	95%	98%
	Said s/he used disordinal interaction?	16	2	5	9		88%	95%	88%	82%
	Used disordinal interaction properly?	11	2	5	4		92%	95%	88%	92%

Notes. C1, C2, and C3 represent the 3 different feedback conditions. All coding was binary ("yes" or "no"), so each disagreement = + 1.

Table 14. *Incremental Prediction of Change Over Time in Judgment Validity (r_a): Insight (Simpler Model = Random Intercepts, Random Slopes, No Time x Feedback Interaction) (Fall 2009)*

		Fix Eff Ind Diff	Std Err Fix Eff	t-val Fix Eff	t-val 95% (|t| >2)	df simpler model	df more complex model	AIC simpler model	AIC more complex model	BIC simpler model	BIC more complex model	logLik simpler model	logLik more complex mod	Chisq	Chi df	pval (Chisq)	sig
1	TrPerExprAware	0.0108	0.0071	1.52		6	7	-555.15	-535.01	-529.09	-504.97	283.57	274.51	0.00	1	1.00E+00	
2	TrPerSaidUsed	0.0126	0.0078	1.61		6	7	-555.15	-527.79	-529.09	-497.80	283.57	270.89	0.00	1	1.00E+00	
3	TrPerUsedCorrectly	0.0230	0.0114	2.02	!	6	7	-555.15	-516.59	-529.09	-486.86	283.57	265.29	0.00	1	1.00E+00	

Table 15. *Incremental Prediction of Change Over Time in Skill Score Accuracy: Insight (Simpler Model = Random Intercepts, Random Slopes, Time x Feedback Interaction) (Fall 2009)*

		Fix Eff Ind Diff	Std Err Fix Eff	t-val Fix Eff	t-val 95% (|t|>2)	df simpler model	df more complex model	AIC simpler model	AIC more complex model	BIC simpler model	BIC more complex model	logLik simpler model	logLik more complex mod	Chisq	Chi df	pval (Chisq)	sig
1	TrPerExprAware	-0.024	0.062	-0.39		10	11	1701.70	1622.05	1745.12	1669.26	-840.85	-800.03	81.65	1	1.63E-19	**
2	TrPerSaidUsed	0.003	0.068	0.05		10	11	1701.70	1610.00	1745.12	1657.12	-840.85	-794.00	93.70	1	3.67E-22	**
3	TrPerUsedCorrectly	0.040	0.100	0.40		10	11	1701.70	1555.48	1745.12	1602.18	-840.85	-766.74	148.22	1	4.25E-34	**

Notes. Fix Eff = fixed effect; Ind Diff = individual difference; Std Err = standard error; t-val = t-value; df = degreesof freedom; AIC = Akaike information criterion, BIC = Bayesian information criterion; logLik = log likelihood; chi or chisq = chi-square χ^2; pval = probability value; simpler model = regression model that best fits the data without the individual difference variable included; more complex model = simpler model with the individual difference included; ! = absolute value for the t-value is greater than 2 (±1.96 = z-score for 2-tailed z-test with $\alpha = 0.05$); ** = p-value < 0.01; * = p-value < 0.05.

Table 16. *Insight: Dimensions Coded Based on Narrative Self-Reports (Fall 2009)*

When Insight Is Achieved	Aware of Disordinal Interaction					
	All Reversals Ignored		Any Reversal → Insight Indeterminate		Any Reversal → No Insight	
	N	%	*N*	%	*N*	%
Never	42	34%	42	45%	74	59%
After block 1	52	42%	30	32%	30	24%
After block 2	8	6%	3	3%	3	2%
After block 3	13	10%	9	10%	9	7%
After block 4	10	8%	9	10%	9	7%
Total	125	100%	93	100%	125	100%

When Insight Is Achieved	Used Disordinal Interaction					
	All Reversals Ignored		Any Reversal → Insight Indeterminate		Any Reversal → No Insight	
	N	%	*N*	%	*N*	%
Never	44	35%	43	51%	83	67%
After block 1	49	40%	19	23%	19	15%
After block 2	11	9%	6	7%	6	5%
After block 3	16	13%	12	14%	12	10%
After block 4	4	3%	4	5%	4	3%
Total	124	100%	84	100%	124	100%

When Insight Is Achieved	Properly Used Disordinal Interaction					
	All Reversals Ignored		Any Reversal → Insight Indeterminate		Any Reversal → No Insight	
	N	%	*N*	%	*N*	%
Never	75	82%	75	87%	80	88%
After block 1	11	12%	7	8%	7	8%
After block 2	2	2%	1	1%	1	1%
After block 3	0	0%	0	0%	0	0%
After block 4	3	3%	3	3%	3	3%
Total	91	100%	86	100%	91	100%

Notes. "All Reversals Ignored" = insight is achieved even if a reversal later occurs; "Any Reversal → Insight Indeterminate" = subject is removed from the analysis if there is any reversal; "Any Reversal → No Insight" = subject is considered to have not gained insight if there is any reversal; *N* = number of subjects; % = percentage of subjects (based on *N*s)

Table 17. *Mechanical Versus Clinical Criterion-Related Accuracy Analyses (Spring 2010)*

	Mechanical		Clinical					z (Sig. if > 1.96)
	Accuracy	Zp of Accuracy	Feedback Group	Mean Estimate	95% CI of Mean Estimate	Zr of Mean Estimate	N of Mean Estimate	
Criterion-Related Validity (R_e or r_a)	0.367	0.385	Group 1	0.433	[0.164, 0.642]	0.464	46	0.54
	0.367	0.385	Group 2	0.417	[0.151, 0.626]	0.444	48	0.41
	0.367	0.385	Group 3	0.527	[0.283, 0.707]	0.586	47	1.38
Skill Score	0.134	--	Group 1	-0.364	--	--	--	--
	0.134	--	Group 2	-0.465	--	--	--	--
	0.134	--	Group 3	-0.244	--	--	--	--

Notes. Depicted are only the data necessary to reject (successfully or not) null hypotheses that clinical validity or accuracy is larger than mechanical validity or accuracy. Therefore, if a sample-level estimate of clinical validity or accuracy is smaller than mechanical clinical validity or accuracy, then no additional data are reported for the sample-level estimate. Zp = z-score of mechanical accuracy; Zr = z-score of clinical mean estimate; z = z-score calculated to test if Zp and Zr are significantly different from each other (see Cohen, 2001, p. 268); 95% CI of Mean Difference = 95% confidence interval for clinical mean estimate; N of mean estimate = sample size for clinical mean estimate (used in z-test).

Table 18. *Changes Over Time in Accuracy and the Determinants of Accuracy (Spring 2010)*

	Fisher r_a			*Skill Score*			*Fisher C*			*Fisher r_z*		
	Estimate	p-value or 95% Confidence Interval	N_{obs} (N_{sub})	Estimate	p-value or 95% Confidence Interval	N_{obs} (N_{sub})	Estimate	p-value or 95% Confidence Interval	N_{obs} (N_{sub})	Estimate	p-value or 95% Confidence Interval	N_{obs} (N_{sub})
Step 1: Fit of Random Intercepts												
ICC:	0	--	705 (141)	0.03	--	705 (141)	0.18	--	705 (141)	0.17	--	705 (141)
Likelihood Ratio χ^2 (where df = 1):	$3.845*10^{-7}$	0.9995	705 (141)	0.754	0.385	705 (141)	46.12	< 0.0001	705 (141)	44.734	< 0.0001	705 (141)
Step 2: Fit of Random Slopes												
Log likelihood χ^2 (where df = 2):	--	--	--	--	--	--	13.761	0.001	705 (141)	12.588	0.001847	705 (141)
Step 3: Fit of Interaction (Feedback Condition x Time)												
Log likelihood χ^2 or F (where df = 4):	2.23	0.064	705 (141)	1.1	0.355	705 (141)	6.724	0.151	705 (141)	6.12	0.1905	705 (141)
Step 4: Slope(s) (unstandardized *γ* or *b* values) of Growth Curve(s)												
Overall (Fixed Effect):	0.034	[0.016, 0.052]	705 (141)	0.063	[-0.037, 0.114]	705 (141)	0.074	[0.053, 0.094]	705 (141)	0.066	[0.049, 0.084]	705 (141)
Feedback Group 1:	0.03	[-0.001, 0.062]	230 (46)	--	--	--	--	--	--	--	--	--
Feedback Group 2:	0.031	[0.001, 0.062]	240 (48)	--	--	--	--	--	--	--	--	--
Feedback Group 3:	0.04	[0.009, 0.071]	235 (47)	--	--	--	--	--	--	--	--	--

Notes. N_{obs} = number of observations based upon which the estimate was calculated; N_{sub} = number of subjects based upon which the estimate was calculated; df = degrees of freedom

Table 18 – *cont'd.*

	Fisher G			*Fisher R_s*			*C_{xj} (Cwtx1x2_s)*		
	Estimate	p-value or 95% Confidence Interval	N_{obs} (N_{sub})	Estimate	p-value or 95% Confidence Interval	N_{obs} (N_{sub})	Estimate	p-value or 95% Confidence Interval	N_{obs} (N_{sub})
<u>Step 1: Fit of Random Intercepts</u>									
ICC:	0	--	705 (141)	0.27	--	705 (141)	0.26	--	705 (141)
Likelihood Ratio χ^2 (where df = 1):	6.78 $*10^{-7}$	0.9993	705 (141)	99.773	< 0.0001	705 (141)	81.981	< 0.0001	705 (141)
<u>Step 2: Fit of Random Slopes</u>									
Log likelihood χ^2 (where df = 2):	--	--	--	57.977	$2.573*10^{-13}$	705 (141)	74.458	$2.2*10^{-16}$	705 (141)
<u>Step 3: Fit of Interaction (Feedback Condition x Time)</u>									
Log likelihood χ^2 or F (where df = 4):	0.618	0.65	705 (141)	11.608	0.021	705 (141)	12.473	0.014	705 (141)
<u>Step 4: Slope(s) (unstandardized *γ* or *b* values) of Growth Curve(s)</u>									
Overall (Fixed Effect):	-0.079	[-0.141, -0.018]	705 (141)	-0.162	[-0.202, -0.123]	705 (141)	0.019	[0.012, 0.026]	705 (141)
Feedback Group 1:	--	--	--	-0.151	[-0.229, -0.073]	230 (46)	0.013	[0.0001, 0.026]	230 (46)
Feedback Group 2:	--	--	--	-0.108	[-0.184, -0.031]	240 (48)	0.018	[0.005, 0.031]	240 (48)
Feedback Group 3:	--	--	--	-0.23	[-0.307, -0.153]	235 (47)	0.026	[0.013, 0.039]	235 (47)

<u>Notes</u>. N_{obs} = number of observations based upon which the estimate was calculated; N_{sub} = number of subjects based upon which the estimate was calculated; df = degrees of freedom

Table 19. *Correlations Among Second-Level (Between-Persons) Variables (Spring 2010)*

	Fisherra	Fisherra1	Fisherra2	Fisherra3	Fisherra4	Fisherra5	SS	SS1	SS2	SS3	SS4	SS5
Fisherra	1**	0.22**	0.4**	0.45**	0.52**	0.41**	0.67**	0.26**	0.37**	0.38**	0.39**	0.27**
Fisherra1	0.22**	1**	-0.26**	-0.17*	-0.08	-0.1	-0.04	0.58**	-0.2*	-0.15	-0.09	-0.19*
Fisherra2	0.4**	-0.26**	1**	-0.01	0.11	0	0.26**	-0.12	0.66**	0	0.2*	0.03
Fisherra3	0.45**	-0.17*	-0.01	1**	0.13	0.05	0.38**	-0.05	0.1	0.67**	0.09	0.18*
Fisherra4	0.52**	-0.08	0.11	0.13	1**	0.01	0.38**	0.05	0.13	0.21*	0.61**	0.05
Fisherra5	0.41**	-0.1	0	0.05	0.01	1**	0.26**	-0.06	-0.01	0.13	-0.04	0.33**
SS	0.67**	-0.04	0.26**	0.38**	0.38**	0.26**	1**	0.23**	0.38**	0.57**	0.47**	0.68**
SS1	0.26**	0.58**	-0.12	-0.05	0.05	-0.06	0.23**	1**	-0.1	0	-0.07	-0.1
SS2	0.37**	-0.2*	0.66**	0.1	0.13	-0.01	0.38**	-0.1	1**	0.15	0.19*	0
SS3	0.38**	-0.15	0	0.67**	0.21*	0.13	0.57**	0	0.15	1**	0.14	0.21*
SS4	0.39**	-0.09	0.2*	0.09	0.61**	-0.04	0.47**	-0.07	0.19*	0.14	1**	0.05
SS5	0.27**	-0.19*	0.03	0.18*	0.05	0.33**	0.68**	-0.1	0	0.21*	0.05	1**
FisherRslin	-0.35**	-0.04	-0.15	-0.18*	-0.28**	-0.07	-0.17*	0.02	-0.11	-0.2*	-0.27**	0.02
FisherRslin1	0.23**	-0.11	0.14	0.1	0.11	0.31**	0.21*	0	0.02	0.16	0.08	0.15
FisherRslin2	-0.16	0.11	-0.09	-0.27**	-0.13	0.06	-0.13	0.09	-0.11	-0.13	-0.07	-0.1
FisherRslin3	-0.26**	-0.03	0.05	-0.3**	-0.25**	-0.01	-0.14	-0.02	0.03	-0.21*	-0.19*	-0.01
FisherRslin4	-0.29**	0.04	-0.15	-0.11	-0.34**	-0.06	-0.21*	0.02	-0.05	-0.15	-0.31**	-0.07
FisherRslin5	-0.39**	-0.1	-0.16	-0.11	-0.22**	-0.22**	-0.29**	0.02	-0.05	-0.18*	-0.29**	-0.19*

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	Fisherra	Fisherra1	Fisherra2	Fisherra3	Fisherra4	Fisherra5	SS	SS1	SS2	SS3	SS4	SS5
FisherG	0.08	0.05	0.11	0	-0.03	0.1	0.13	0.13	-0.02	0.07	-0.04	0.11
FisherG1	0.06	0.59**	-0.2*	-0.17*	0.01	-0.05	-0.12	0.35**	-0.25**	-0.11	-0.07	-0.18*
FisherG2	0.07	-0.15	0.54**	-0.23**	0.06	-0.08	0.02	-0.16	0.32**	-0.11	0.14	-0.06
FisherG3	0.12	-0.14	-0.12	0.55**	-0.11	0.07	0.12	-0.03	-0.02	0.29**	-0.07	0.12
FisherG4	0.13	-0.15	-0.03	0.19*	0.42**	-0.13	0.09	-0.06	0.07	0.12	0.24**	-0.05
FisherG5	0.06	-0.18*	-0.07	-0.04	-0.19*	0.64**	0.01	-0.07	-0.05	-0.02	-0.17*	0.14
FisherCstat	0.93**	0.18*	0.38**	0.42**	0.5**	0.35**	0.6**	0.21*	0.34**	0.34**	0.42**	0.23**
FisherCstat1	0.21*	0.56**	-0.01	0.02	-0.13	-0.12	0.02	0.3**	-0.02	0	-0.01	-0.13
FisherCstat2	0.59**	-0.11	0.63**	0.3**	0.25**	0.11	0.48**	0.06	0.5**	0.23**	0.23**	0.21*
FisherCstat3	0.61**	0.01	0.14	0.61**	0.43**	0.08	0.43**	0.13	0.13	0.35**	0.29**	0.17*
FisherCstat4	0.65**	-0.01	0.26**	0.21*	0.67**	0.31**	0.44**	0.01	0.18*	0.25**	0.45**	0.19*
FisherCstat5	0.55**	0.06	0.18*	0.17*	0.27**	0.65**	0.23**	0	0.1	0.2*	0.17*	0.07
Fisherrz	0.93**	0.18*	0.38**	0.42**	0.5**	0.35**	0.6**	0.21*	0.34**	0.34**	0.42**	0.23**
Fisherrz1	0.22**	0.49**	0.02	0.04	-0.12	-0.1	0.06	0.27**	0	0.03	0.01	-0.09
Fisherrz2	0.59**	-0.09	0.52**	0.33**	0.27**	0.13	0.49**	0.07	0.47**	0.26**	0.23**	0.22**
Fisherrz3	0.61**	0.03	0.15	0.53**	0.45**	0.1	0.43**	0.15	0.13	0.29**	0.3**	0.19*
Fisherrz4	0.62**	0	0.3**	0.2*	0.54**	0.34**	0.41**	-0.01	0.19*	0.24**	0.39**	0.19*
Fisherrz5	0.57**	0.1	0.22**	0.17*	0.31**	0.52**	0.23**	0.03	0.15	0.18*	0.21*	0.02

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	Fisherra	Fisherra1	Fisherra2	Fisherra3	Fisherra4	Fisherra5	SS	SS1	SS2	SS3	SS4	SS5
Cwtx1x2_s	0.65**	0.09	0.29**	0.24**	0.4**	0.21*	0.4**	0.14	0.26**	0.16	0.27**	0.18*
Cwtx1x2_1_s	0.2*	0.25**	0.09	0.05	0.03	0.02	0.01	0.15	0.02	0.03	-0.08	-0.09
Cwtx1x2_2_s	0.41**	-0.06	0.42**	0.18*	0.16	0.23**	0.33**	0.06	0.29**	0.26**	0.15	0.1
Cwtx1x2_3_s	0.54**	0.04	0.13	0.47**	0.37**	0.2*	0.43**	0.1	0.19*	0.4**	0.3**	0.13
Cwtx1x2_4_s	0.55**	0.12	0.23**	0.15	0.47**	0.25**	0.37**	0.07	0.18*	0.16	0.35**	0.17*
Cwtx1x2_5_s	0.57**	0.09	0.22*	0.12	0.28**	0.55**	0.38**	0.09	0.13	0.19*	0.22**	0.2*
ConfAbsOverallPre	0.04	0.03	0.15	-0.19*	-0.07	0.07	0.08	-0.03	0.22*	-0.01	-0.01	0.04
ConfAbs1	0.03	0.13	0.07	-0.01	-0.18*	0.02	-0.04	0.09	-0.01	-0.02	-0.12	-0.05
ConfAbs2	0.06	0.15	0.03	-0.08	-0.11	0.06	0.08	0.06	-0.09	0.03	-0.06	0.14
ConfAbs3	0.04	0.04	0.06	-0.01	-0.1	0.08	0.05	0.03	-0.01	0.08	-0.09	0.07
ConfAbs4	0.16	0.13	0.09	-0.01	0.05	0.02	0.13	0.08	0.05	0.06	0.06	0.07
ConfAbs5	0.18*	-0.01	0.13	-0.04	0.05	0.18*	0.12	0.1	0.04	0.02	-0.01	0.16
ConfAbsOverallPost	0.15	0.12	0.07	-0.03	-0.02	0.11	0.01	0.01	-0.04	-0.02	-0.12	0.08
ConfRelOverallPre	0.08	-0.05	0.11	-0.05	0.02	0.09	0.07	-0.13	0.12	0.14	0.04	0.01
ConfRel1	-0.01	0.18*	0.01	-0.01	-0.18*	-0.01	-0.02	0.1	-0.06	0.04	-0.16	0.02
ConfRel2	0.07	0.1	0.05	-0.04	-0.12	0.06	0.09	0.04	0.02	0.08	-0.09	0.09
ConfRel3	0.09	-0.01	0.14	0.01	-0.03	0.04	0.09	-0.06	0.07	0.1	0.01	0.06
ConfRel4	0.17*	0.13	0.15	-0.04	0.03	0.04	0.13	0.08	0.09	0.05	0.02	0.06
ConfRel5	0.18*	0.12	0.13	-0.05	-0.04	0.13	0.13	0.13	0.05	0.04	-0.05	0.15
ConfRelOverallPost	0.19*	0.2*	0.11	-0.1	0.01	0.11	0.09	-0.01	0.04	0.01	-0.01	0.09

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	Fisherra	Fisherra1	Fisherra2	Fisherra3	Fisherra4	Fisherra5	SS	SS1	SS2	SS3	SS4	SS5
ACT_COMP_SCR	0.21*	0.11	0.1	0.02	0.08	0.21*	0.38**	-0.05	0.21*	0.17	0.25*	0.24*
ACT_ENGL_SCR	0.18	0.04	0.01	0.02	0.07	0.24*	0.33**	0	0.11	0.18	0.23*	0.2
ACT_ENGWR_SCR	0.19	0	-0.03	0.02	0.19	0.26*	0.3**	-0.03	0	0.15	0.28**	0.24*
ACT_MATH_SCR	0.32**	0.13	0.09	0.17	0.15	0.2	0.39**	0.26*	0.15	0.18	0.2*	0.2*
ACT_READ_SCR	0.1	0.09	0.08	0.01	-0.03	0.07	0.31**	0.05	0.18	0.14	0.18	0.14
ACT_SCIRE_SCR	0.15	0.07	0.02	0.03	0.08	0.14	0.26*	0.07	0.11	0.07	0.14	0.21*
CUM_GPA	0.09	0.16	0.06	0.04	0.02	0.03	0.1	0.12	0	0.14	0.09	-0.06
TOT_ACAD_HOURS	-0.08	-0.13	-0.02	0.22*	-0.19*	-0.02	0.04	-0.13	0.07	0.12	-0.06	0.08
HS_RANK_PCT	0.18	-0.11	0.16	0.07	0.12	0.24*	0.26*	-0.02	0.14	0.16	0.2	0.12
Gender	0.11	0.06	-0.04	0.04	0.02	0.14	0.03	0.06	0.11	0.09	-0.05	-0.07
extroGoldberg	0.09	-0.03	0.07	0.03	0.01	0.07	0.06	-0.09	0.11	-0.02	0	0.09
neurotGoldberg	-0.03	-0.09	-0.08	0.06	-0.03	-0.01	-0.12	-0.1	-0.03	0.02	0	-0.1
intelGoldberg	0.12	-0.11	0.13	0.13	0.02	0.11	0.15	-0.08	0.15	0.06	0.15	0.08
agreeGoldberg	0.04	0.06	0	0.07	-0.04	-0.01	0.07	0.05	-0.08	0.02	0	0.1
conGoldberg	0.1	0.12	0.04	0.1	0.05	-0.02	0.08	0.02	0.03	0.12	0.04	-0.01
pvIPC7	0.15	0.14	0.12	-0.14	0.02	0.17	0.05	0.1	0.03	-0.13	0.06	0.02
nvIPC7	-0.1	-0.06	-0.01	0	-0.02	-0.13	-0.13	-0.12	0.02	-0.06	0.01	-0.14
pemIPC7	0.04	-0.04	0.09	0.02	-0.05	0.06	0.05	0	0.09	-0.04	-0.05	0.07
nemIPC7	-0.02	-0.13	-0.1	0.13	-0.02	0.01	0	-0.08	-0.1	0.02	0.05	0.07
conIPC7	0.11	0.17	-0.02	0.08	0.06	0	0.01	0.11	-0.05	0.09	-0.04	-0.04
agreeIPC7	0.05	0.05	0.03	0.12	0	-0.06	0.01	0.07	-0.02	0.04	-0.06	0.01
cnvIPC7	-0.01	0.15	-0.06	-0.06	-0.15	0.05	-0.09	0.2*	-0.04	-0.07	-0.17	-0.09

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	Fisherra	Fisherra1	Fisherra2	Fisherra3	Fisherra4	Fisherra5	SS	SS1	SS2	SS3	SS4	SS5
ReaRIASEC	-0.04	0.06	-0.05	-0.01	-0.02	-0.06	0.03	0.05	0	0.03	-0.06	0.03
InvRIASEC	-0.09	-0.05	-0.06	-0.07	0	0.06	-0.01	0	-0.05	-0.09	-0.03	0.06
ArtRIASEC	-0.03	-0.14	0.06	0.02	-0.03	0.01	0.09	-0.04	0.08	0.06	0.02	0.09
SocRIASEC	-0.17	-0.09	0.09	-0.01	-0.18*	-0.06	0.02	0	0.02	0.04	-0.03	0.01
EntRIASEC	-0.03	0.02	0.11	-0.15	0.01	-0.02	-0.04	0.04	0.01	-0.06	0.09	-0.12
ConRIASEC	-0.09	0.06	0.03	-0.05	-0.04	-0.12	-0.14	-0.07	-0.07	-0.11	0	-0.11
numpredout	-0.08	-0.1	0.01	-0.14	0.1	-0.01	-0.04	-0.04	0.08	-0.01	0.07	-0.1
numratees	-0.04	-0.12	0.03	-0.03	0.04	0	-0.03	-0.02	0.06	-0.01	0.02	-0.04
lengthtimepred	-0.1	-0.09	-0.04	0.03	-0.13	0.01	-0.02	0	0.05	-0.04	-0.09	0.01
numpredoutformal	-0.18*	-0.12	-0.05	-0.05	-0.09	-0.1	-0.08	-0.02	0.01	-0.11	-0.03	-0.01
numpredouttraining	0.01	-0.07	0.16	-0.04	-0.02	0.03	-0.02	-0.11	0.07	-0.14	0.02	0.04
numpredouttrainingformal	-0.09	-0.08	-0.09	-0.01	-0.06	-0.02	-0.04	0	-0.07	0	-0.01	0.02
lengthtraining	-0.1	-0.15	0.11	-0.04	-0.02	-0.06	-0.08	-0.06	0.07	0.01	0.03	-0.14
amtstats	-0.04	-0.21*	0.11	0.06	0.05	-0.08	0.04	-0.12	0.01	0.06	0.1	0.05
amtdecisionmkg	0.14	-0.1	0.06	0.1	0.1	0.1	0.1	0	0.04	0.09	0.09	0.05
TrPerExprAware	0.25**	0	-0.02	0.14	0.24**	0.23**	0.24**	0.03	0.07	0.26**	0.26**	0.03
TrPerSaidUsed	0.29**	-0.04	-0.03	0.2*	0.21*	0.28**	0.32**	0.03	0.06	0.32**	0.21*	0.16
TrPerUsedCorrectly	0.3**	-0.11	0.04	0.28**	0.33**	0.13	0.26**	-0.09	0.06	0.26**	0.24*	0.18
TrPerSaidUsed3rdVar	-0.03	-0.1	0.14	-0.17*	0.05	0.03	0.05	0	0.09	-0.08	0.1	0.01

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	FisherRslin	FisherRslin1	FisherRslin2	FisherRslin3	FisherRslin4	FisherRslin5	FisherG	FisherG1	FisherG2	FisherG3	FisherG4	FisherG5
Fisherra	-0.35**	0.23**	-0.16	-0.26**	-0.29**	-0.39**	0.08	0.06	0.07	0.12	0.13	0.06
Fisherra1	-0.04	-0.11	0.11	-0.03	0.04	-0.1	0.05	0.59**	-0.15	-0.14	-0.15	-0.18*
Fisherra2	-0.15	0.14	-0.09	0.05	-0.15	-0.16	0.11	-0.2*	0.54**	-0.12	-0.03	-0.07
Fisherra3	-0.18*	0.1	-0.27**	-0.3**	-0.11	-0.11	0	-0.17*	-0.23**	0.55**	0.19*	-0.04
Fisherra4	-0.28**	0.11	-0.13	-0.25**	-0.34**	-0.22**	-0.03	0.01	0.06	-0.11	0.42**	-0.19*
Fisherra5	-0.07	0.31**	0.06	-0.01	-0.06	-0.22**	0.1	-0.05	-0.08	0.07	-0.13	0.64**
SS	-0.17*	0.21*	-0.13	-0.14	-0.21*	-0.29**	0.13	-0.12	0.02	0.12	0.09	0.01
SS1	0.02	0	0.09	-0.02	0.02	0.02	0.13	0.35**	-0.16	-0.03	-0.06	-0.07
SS2	-0.11	0.02	-0.11	0.03	-0.05	-0.05	-0.02	-0.25**	0.32**	-0.02	0.07	-0.05
SS3	-0.2*	0.16	-0.13	-0.21*	-0.15	-0.18*	0.07	-0.11	-0.11	0.29**	0.12	-0.02
SS4	-0.27**	0.08	-0.07	-0.19*	-0.31**	-0.29**	-0.04	-0.07	0.14	-0.07	0.24**	-0.17*
SS5	0.02	0.15	-0.1	-0.01	-0.07	-0.19*	0.11	-0.18*	-0.06	0.12	-0.05	0.14
FisherRslin	1**	0.33**	0.46**	0.73**	0.81**	0.63**	0.12	0.04	-0.02	0.05	0	0.17*
FisherRslin1	0.33**	1**	0.37**	0.33**	0.18*	-0.08	0.13	-0.01	0.09	0.09	0.04	0.26**
FisherRslin2	0.46**	0.37**	1**	0.43**	0.36**	0.06	0.11	0.24**	0.1	-0.07	0	0.05
FisherRslin3	0.73**	0.33**	0.43**	1**	0.69**	0.41**	0.05	0	0.17*	-0.03	-0.11	0.11
FisherRslin4	0.81**	0.18*	0.36**	0.69**	1**	0.49**	0.1	0.01	-0.08	0.04	-0.05	0.16
FisherRslin5	0.63**	-0.08	0.06	0.41**	0.49**	1**	0.09	-0.07	-0.07	0.12	0	0.15
FisherG	0.12	0.13	0.11	0.05	0.1	0.09	1**	0.23**	0.14	0.09	-0.09	0.16
FisherG1	0.04	-0.01	0.24**	0	0.01	-0.07	0.23**	1**	-0.14	-0.16	-0.05	-0.1
FisherG2	-0.02	0.09	0.1	0.17*	-0.08	-0.07	0.14	-0.14	1**	-0.21*	-0.09	-0.03
FisherG3	0.05	0.09	-0.07	-0.03	0.04	0.12	0.09	-0.16	-0.21*	1**	0.1	0.11
FisherG4	0	0.04	0	-0.11	-0.05	0	-0.09	-0.05	-0.09	0.1	1**	-0.06
FisherG5	0.17*	0.26**	0.05	0.11	0.16	0.15	0.16	-0.1	-0.03	0.11	-0.06	1**

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	FisherRslin	FisherRslin1	FisherRslin2	FisherRslin3	FisherRslin4	FisherRslin5	FisherG	FisherG1	FisherG2	FisherG3	FisherG4	FisherG5
FisherCstat	-0.46**	0.17*	-0.21*	-0.33**	-0.37**	-0.47**	-0.07	0.01	0.07	0.07	0.12	0
FisherCstat1	-0.07	-0.14	-0.09	-0.02	0	0.03	-0.14	0.04	0.07	0.03	-0.1	-0.07
FisherCstat2	-0.19*	0.15	-0.31**	-0.11	-0.16	-0.19*	0.03	-0.27**	0.14	0.03	0.03	-0.06
FisherCstat3	-0.27**	0.08	-0.28**	-0.34**	-0.22*	-0.19*	-0.07	-0.07	-0.07	0.06	0.2*	-0.09
FisherCstat4	-0.42**	0.11	-0.16	-0.31**	-0.43**	-0.4**	-0.01	0.02	0.07	-0.07	0.07	-0.03
FisherCstat5	-0.22**	0.31**	0.09	-0.13	-0.21*	-0.39**	0.05	0.06	0.01	0.01	0.1	0.2*
Fisherrz	-0.46**	0.17*	-0.21*	-0.32**	-0.37**	-0.47**	-0.07	0.01	0.07	0.07	0.12	0
Fisherrz1	-0.07	-0.12	-0.1	-0.02	-0.02	0.03	-0.15	0.01	0.09	0.03	-0.09	-0.05
Fisherrz2	-0.2*	0.13	-0.32**	-0.13	-0.15	-0.18*	0	-0.25**	0.08	0.07	0.04	-0.04
Fisherrz3	-0.26**	0.1	-0.25**	-0.33**	-0.21*	-0.2*	-0.06	-0.07	-0.05	0	0.2*	-0.08
Fisherrz4	-0.43**	0.09	-0.14	-0.29**	-0.44**	-0.41**	0	0.02	0.07	-0.05	0	0
Fisherrz5	-0.21*	0.3**	0.11	-0.11	-0.19*	-0.37**	0.06	0.08	0.06	0	0.15	0.1
Cwtx1x2_s	-0.12	0.25**	-0.12	-0.12	-0.2*	-0.25**	-0.11	-0.04	0.1	0.01	0.29**	0.08
Cwtx1x2_1_s	0.03	0.13	0.04	0.02	0.02	0.09	-0.12	0.01	-0.05	0.05	-0.05	-0.03
Cwtx1x2_2_s	-0.21*	0.1	-0.26**	-0.12	-0.15	-0.13	0.07	-0.12	-0.08	-0.1	-0.09	0.03
Cwtx1x2_3_s	-0.25**	0.12	-0.09	-0.26**	-0.2*	-0.28**	0.01	0.03	-0.11	-0.03	0.16	-0.05
Cwtx1x2_4_s	-0.28**	0.23**	0.01	-0.17*	-0.29**	-0.31**	0.02	0.12	0.05	-0.1	-0.09	-0.07
Cwtx1x2_5_s	-0.28**	0.25**	0.07	-0.13	-0.27**	-0.39**	0.1	0.04	0.03	-0.01	-0.03	0.14

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	FisherRslin	FisherRslin1	FisherRslin2	FisherRslin3	FisherRslin4	FisherRslin5	FisherG	FisherG1	FisherG2	FisherG3	FisherG4	FisherG5
ConfAbsOverallPre	0.04	0.1	0.06	0.1	0.01	0.06	0.21*	0.1	0.07	-0.16	-0.17	-0.02
ConfAbs1	0.01	0.01	0	0.04	-0.02	0.03	0.11	0.13	0.05	-0.06	-0.26**	-0.02
ConfAbs2	-0.14	-0.08	-0.08	-0.11	-0.13	-0.09	0.05	0.08	0	-0.07	-0.34**	-0.1
ConfAbs3	-0.04	-0.12	-0.07	0.09	-0.02	-0.02	0.07	0.09	0.06	-0.01	-0.21*	-0.04
ConfAbs4	-0.14	-0.05	-0.03	0.01	-0.02	-0.19*	0.01	0.17*	0.07	-0.06	-0.08	-0.14
ConfAbs5	-0.11	-0.01	-0.02	0.01	-0.06	-0.06	-0.01	0.02	0.09	-0.07	-0.16	-0.04
ConfAbsOverallPost	-0.1	-0.07	-0.14	-0.02	-0.02	-0.03	0.06	0.07	0	0	-0.16	0
ConfRelOverallPre	-0.04	0.08	-0.08	-0.04	-0.02	0	0.1	-0.02	0.04	-0.11	-0.16	0.03
ConfRel1	0	-0.03	-0.07	-0.01	0	0.06	0.1	0.15	0	-0.09	-0.27**	-0.08
ConfRel2	-0.12	-0.08	-0.17*	-0.11	-0.08	0	0.01	0.04	-0.06	-0.06	-0.23**	-0.04
ConfRel3	-0.01	-0.05	-0.07	0.05	0.01	0	0.03	0.04	0.09	-0.03	-0.11	-0.08
ConfRel4	-0.11	-0.08	-0.08	-0.01	-0.04	-0.09	-0.05	0.11	0.08	-0.14	-0.07	-0.15
ConfRel5	-0.13	-0.05	-0.05	-0.02	-0.05	-0.09	-0.02	0.09	0.09	-0.1	-0.16	-0.1
ConfRelOverallPost	-0.07	-0.05	-0.1	0.02	0.01	-0.07	0.03	0.12	0.1	-0.13	-0.16	-0.05
ACT_COMP_SCR	0.09	0.37**	0.14	0.07	0.01	-0.23*	0.07	0.14	0.04	-0.16	-0.12	-0.04
ACT_ENGL_SCR	0.11	0.38**	0.12	0.12	0.02	-0.16	0.08	0.14	-0.03	-0.1	-0.11	0.05
ACT_ENGWR_SCR	0.13	0.43**	0.15	0.12	0	-0.16	0.06	0.14	-0.06	-0.11	0.03	0.06
ACT_MATH_SCR	-0.12	0.24*	0	-0.12	-0.15	-0.28**	0.11	0.17	-0.06	-0.04	-0.12	-0.11
ACT_READ_SCR	0.19	0.33**	0.17	0.17	0.15	-0.08	0.1	0.08	0.06	-0.14	0	-0.03
ACT_SCIRE_SCR	0.1	0.33**	0.07	0.09	0.02	-0.15	-0.04	0.12	0	-0.09	-0.06	-0.11
CUM_GPA	-0.06	0.06	0.11	0.05	-0.04	-0.12	0.08	0.06	-0.09	-0.08	-0.08	-0.09
TOT_ACAD_HOURS	0.15	-0.05	0.09	0.18*	0.15	0.11	0.06	-0.08	0.01	0.08	0.06	-0.04
HS_RANK_PCT	-0.2	0.16	0.07	-0.17	-0.16	-0.44**	0.04	-0.06	0.11	-0.12	0.05	0.09
Gender	-0.11	-0.07	-0.12	-0.17	-0.14	0.01	-0.03	0.01	-0.07	0	-0.2*	0.06

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	FisherRslin	FisherRslin1	FisherRslin2	FisherRslin3	FisherRslin4	FisherRslin5	FisherG	FisherG1	FisherG2	FisherG3	FisherG4	FisherG5
extroGoldberg	-0.13	-0.15	-0.11	-0.05	-0.11	0.01	0.06	-0.06	-0.01	0.08	0.01	0.01
neurotGoldberg	0.03	0.06	0.11	-0.01	0.03	0	-0.03	-0.07	0.06	0.1	0.01	0.03
intelGoldberg	0	-0.01	0.07	0.01	0.05	0	0.12	-0.11	-0.02	0.02	-0.02	-0.01
agreeGoldberg	-0.02	0.06	-0.01	-0.01	0.01	-0.04	0.17	0.07	-0.1	0.08	0.06	0.01
conGoldberg	-0.12	-0.06	0.04	0.03	-0.03	-0.14	-0.04	0.03	0	-0.03	0.13	-0.09
pvIPC7	-0.11	-0.01	-0.09	-0.05	-0.14	-0.07	0.02	0.19*	0.03	-0.02	-0.08	0.14
nvIPC7	-0.04	-0.09	-0.07	-0.15	-0.06	0.08	-0.02	-0.07	0.01	0.01	0.01	-0.08
pemIPC7	-0.03	-0.13	-0.11	0	-0.07	0.05	0.05	-0.05	-0.04	0.03	0.04	0.05
nemIPC7	-0.1	0.07	0.01	-0.08	-0.12	-0.15	-0.09	-0.09	0	0.03	-0.04	-0.03
conIPC7	-0.09	0.01	0.07	0	-0.04	-0.13	-0.05	0.05	0.03	-0.04	0.1	-0.06
agreeIPC7	0.13	0.14	0.02	0.04	0.17*	0.04	0.08	0.02	-0.09	0.1	0.03	0.05
cnvIPC7	0.02	-0.01	-0.11	0.06	0.05	0.04	-0.13	-0.01	-0.04	-0.01	-0.16	0.14
ReaRIASEC	0.08	-0.1	-0.15	-0.03	0.02	0.14	-0.16	0.02	-0.11	-0.02	0.02	-0.08
InvRIASEC	0.14	-0.06	-0.05	0.07	0.08	0.13	0.1	-0.02	0	0.04	0	0.1
ArtRIASEC	0.12	0	0	0.04	0.06	0.02	0	-0.05	0.04	0.12	0.06	-0.05
SocRIASEC	0.07	-0.04	-0.09	0.15	0.12	0.05	0.04	-0.09	-0.02	0.13	-0.12	-0.07
EntRIASEC	-0.05	-0.03	-0.11	-0.07	-0.05	-0.04	-0.12	0.01	0.03	-0.03	-0.07	-0.07
ConRIASEC	-0.08	-0.07	-0.09	-0.13	-0.07	-0.07	-0.22*	0	0.07	-0.12	-0.03	-0.16

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	FisherRslin	FisherRslin1	FisherRslin2	FisherRslin3	FisherRslin4	FisherRslin5	FisherG	FisherG1	FisherG2	FisherG3	FisherG4	FisherG5
numpredout	-0.08	-0.07	-0.17*	-0.07	-0.05	0.06	0.04	-0.11	-0.08	-0.2*	0.02	-0.03
numratees	-0.02	-0.1	-0.22**	-0.04	-0.03	0.1	0.03	-0.07	0	0.01	0.08	0.04
lengthtimepred	0.27**	-0.05	-0.01	0.25**	0.28**	0.32**	0.04	-0.04	-0.01	0.01	-0.05	0.1
numpredoutformal	0	-0.16	-0.11	-0.05	0.04	0.08	-0.17*	-0.08	-0.09	0.02	0.15	-0.01
numpredouttraining	0.04	0.03	0.05	0.03	-0.01	0.05	-0.05	-0.04	0.01	0	0.13	0.06
numpredouttrainingformal	-0.09	-0.15	-0.02	-0.1	-0.1	0	-0.1	-0.05	-0.16	0.15	0.06	0.01
lengthtraining	0.03	-0.03	-0.04	0.03	0	0.04	-0.03	-0.08	-0.02	-0.05	0.1	0.02
amtstats	-0.1	0.02	0.1	-0.06	0.02	-0.09	0.09	-0.19*	0.07	-0.01	0.07	-0.12
amtdecisionmkg	-0.1	0.01	0.14	-0.09	-0.07	-0.15	-0.05	-0.09	0.05	0.1	0.03	-0.06
TrPerExprAware	-0.29**	0.09	-0.1	-0.17	-0.28**	-0.3**	-0.12	-0.02	0.01	-0.07	-0.02	0.04
TrPerSaidUsed	-0.21*	0.16	-0.1	-0.17*	-0.2*	-0.27**	-0.06	-0.09	-0.04	-0.02	0.01	0.07
TrPerUsedCorrectly	-0.25**	0.09	-0.31**	-0.26**	-0.23*	-0.21*	-0.1	-0.14	-0.02	0	0	-0.04
TrPerSaidUsed3rdVar	-0.06	-0.05	0.05	-0.02	-0.17*	-0.09	0	-0.09	0.22*	-0.09	-0.01	-0.02

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	FisherCstat	FisherCstat1	FisherCstat2	FisherCstat3	FisherCstat4	FisherCstat5	Fisherrz	Fisherrz1	Fisherrz2	Fisherrz3	Fisherrz4	Fisherrz5
Fisherra	0.93**	0.21*	0.59**	0.61**	0.65**	0.55**	0.93**	0.22**	0.59**	0.61**	0.62**	0.57**
Fisherra1	0.18*	0.56**	-0.11	0.01	-0.01	0.06	0.18*	0.49**	-0.09	0.03	0	0.1
Fisherra2	0.38**	-0.01	0.63**	0.14	0.26**	0.18*	0.38**	0.02	0.52**	0.15	0.3**	0.22**
Fisherra3	0.42**	0.02	0.3**	0.61**	0.21*	0.17*	0.42**	0.04	0.33**	0.53**	0.2*	0.17*
Fisherra4	0.5**	-0.13	0.25**	0.43**	0.67**	0.27**	0.5**	-0.12	0.27**	0.45**	0.54**	0.31**
Fisherra5	0.35**	-0.12	0.11	0.08	0.31**	0.65**	0.35**	-0.1	0.13	0.1	0.34**	0.52**
SS	0.6**	0.02	0.48**	0.43**	0.44**	0.23**	0.6**	0.06	0.49**	0.43**	0.41**	0.23**
SS1	0.21*	0.3**	0.06	0.13	0.01	0	0.21*	0.27**	0.07	0.15	-0.01	0.03
SS2	0.34**	-0.02	0.5**	0.13	0.18*	0.1	0.34**	0	0.47**	0.13	0.19*	0.15
SS3	0.34**	0	0.23**	0.35**	0.25**	0.2*	0.34**	0.03	0.26**	0.29**	0.24**	0.18*
SS4	0.42**	-0.01	0.23**	0.29**	0.45**	0.17*	0.42**	0.01	0.23**	0.3**	0.39**	0.21*
SS5	0.23**	-0.13	0.21*	0.17*	0.19*	0.07	0.23**	-0.09	0.22**	0.19*	0.19*	0.02
FisherRslin	-0.46**	-0.07	-0.19*	-0.27**	-0.42**	-0.22**	-0.46**	-0.07	-0.2*	-0.26**	-0.43**	-0.21*
FisherRslin1	0.17*	-0.14	0.15	0.08	0.11	0.31**	0.17*	-0.12	0.13	0.1	0.09	0.3**
FisherRslin2	-0.21*	-0.09	-0.31**	-0.28**	-0.16	0.09	-0.21*	-0.1	-0.32**	-0.25**	-0.14	0.11
FisherRslin3	-0.33**	-0.02	-0.11	-0.34**	-0.31**	-0.13	-0.32**	-0.02	-0.13	-0.33**	-0.29**	-0.11
FisherRslin4	-0.37**	0	-0.16	-0.22*	-0.43**	-0.21*	-0.37**	-0.02	-0.15	-0.21*	-0.44**	-0.19*
FisherRslin5	-0.47**	0.03	-0.19*	-0.19*	-0.4**	-0.39**	-0.47**	0.03	-0.18*	-0.2*	-0.41**	-0.37**
FisherG	-0.07	-0.14	0.03	-0.07	-0.01	0.05	-0.07	-0.15	0	-0.06	0	0.06
FisherG1	0.01	0.04	-0.27**	-0.07	0.02	0.06	0.01	0.01	-0.25**	-0.07	0.02	0.08
FisherG2	0.07	0.07	0.14	-0.07	0.07	0.01	0.07	0.09	0.08	-0.05	0.07	0.06
FisherG3	0.07	0.03	0.03	0.06	-0.07	0.01	0.07	0.03	0.07	0	-0.05	0
FisherG4	0.12	-0.1	0.03	0.2*	0.07	0.1	0.12	-0.09	0.04	0.2*	0	0.15
FisherG5	0	-0.07	-0.06	-0.09	-0.03	0.2*	0	-0.05	-0.04	-0.08	0	0.1

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	FisherCstat	FisherCstat1	FisherCstat2	FisherCstat3	FisherCstat4	FisherCstat5	Fisherrz	Fisherrz1	Fisherrz2	Fisherrz3	Fisherrz4	Fisherrz5
FisherCstat	1**	0.25**	0.62**	0.65**	0.7**	0.56**	1**	0.26**	0.62**	0.65**	0.67**	0.58**
FisherCstat1	0.25**	1**	0.11	0.07	-0.11	-0.04	0.24**	0.99**	0.1	0.05	-0.09	0.01
FisherCstat2	0.62**	0.11	1**	0.44**	0.36**	0.19*	0.62**	0.14	0.98**	0.43**	0.34**	0.2*
FisherCstat3	0.65**	0.07	0.44**	1**	0.39**	0.23**	0.65**	0.08	0.44**	0.98**	0.33**	0.23**
FisherCstat4	0.7**	-0.11	0.36**	0.39**	1**	0.48**	0.7**	-0.1	0.36**	0.41**	0.97**	0.47**
FisherCstat5	0.56**	-0.04	0.19*	0.23**	0.48**	1**	0.56**	-0.05	0.18*	0.24**	0.48**	0.97**
Fisherrz	1**	0.24**	0.62**	0.65**	0.7**	0.56**	1**	0.25**	0.62**	0.65**	0.67**	0.58**
Fisherrz1	0.26**	0.99**	0.14	0.08	-0.1	-0.05	0.25**	1**	0.13	0.06	-0.09	0
Fisherrz2	0.62**	0.1	0.98**	0.44**	0.36**	0.18*	0.62**	0.13	1**	0.43**	0.34**	0.18*
Fisherrz3	0.65**	0.05	0.43**	0.98**	0.41**	0.24**	0.65**	0.06	0.43**	1**	0.33**	0.24**
Fisherrz4	0.67**	-0.09	0.34**	0.33**	0.97**	0.48**	0.67**	-0.09	0.34**	0.33**	1**	0.46**
Fisherrz5	0.58**	0.01	0.2*	0.23**	0.47**	0.97**	0.58**	0	0.18*	0.24**	0.46**	1**
Cwtx1x2_s	0.7**	0.16	0.47**	0.48**	0.48**	0.41**	0.7**	0.18*	0.48**	0.48**	0.44**	0.42**
Cwtx1x2_1_s	0.17*	0.26**	0.08	0.06	0.03	0.14	0.17*	0.23**	0.06	0.03	0.06	0.14
Cwtx1x2_2_s	0.41**	0.11	0.52**	0.27**	0.27**	0.24**	0.4**	0.13	0.48**	0.25**	0.3**	0.21*
Cwtx1x2_3_s	0.53**	-0.02	0.36**	0.59**	0.41**	0.3**	0.53**	-0.01	0.38**	0.57**	0.37**	0.31**
Cwtx1x2_4_s	0.54**	0.01	0.23**	0.37**	0.55**	0.39**	0.54**	0.02	0.22**	0.41**	0.5**	0.4**
Cwtx1x2_5_s	0.54**	0.01	0.23**	0.22**	0.44**	0.67**	0.54**	0	0.2*	0.24**	0.45**	0.66**

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	FisherCstat	FisherCstat1	FisherCstat2	FisherCstat3	FisherCstat4	FisherCstat5	Fisherrz	Fisherrz1	Fisherrz2	Fisherrz3	Fisherrz4	Fisherrz5
ConfAbsOverallPre	0.06	-0.14	0.13	-0.12	0.11	0.08	0.06	-0.14	0.12	-0.11	0.14	0.08
ConfAbs1	0.05	0.07	0.06	0.02	-0.01	0.06	0.05	0.07	0.04	0	0.03	0.05
ConfAbs2	0.11	0.1	0.03	-0.03	0.1	0.11	0.11	0.1	0.01	-0.04	0.14	0.11
ConfAbs3	0.06	0.09	0.05	0	0.07	0.02	0.06	0.1	0.05	-0.03	0.09	0.01
ConfAbs4	0.18*	0.01	0.09	0.08	0.17	0.14	0.18*	0.01	0.08	0.09	0.15	0.16
ConfAbs5	0.18*	-0.02	0.12	0.04	0.17*	0.21*	0.18*	-0.01	0.11	0.04	0.18*	0.19*
ConfAbsOverallPost	0.15	0.07	0.13	0.05	0.16	0.09	0.15	0.07	0.13	0.04	0.19*	0.06
ConfRelOverallPre	0.11	-0.12	0.14	0.03	0.16	0.05	0.11	-0.11	0.14	0.02	0.16	0.03
ConfRel1	-0.01	0.07	0.02	-0.11	-0.02	0.02	-0.01	0.05	0.01	-0.12	0.03	0.01
ConfRel2	0.1	0.09	0.13	-0.05	0.05	0.05	0.1	0.11	0.12	-0.05	0.09	0.05
ConfRel3	0.1	0.04	0.12	0.03	0.06	0	0.1	0.06	0.12	0.02	0.08	0
ConfRel4	0.17	0	0.15	0.09	0.13	0.09	0.16	0	0.14	0.1	0.14	0.09
ConfRel5	0.19*	0.04	0.13	0.03	0.09	0.18*	0.19*	0.04	0.12	0.02	0.12	0.17*
ConfRelOverallPost	0.17*	0.09	0.11	0	0.12	0.11	0.17	0.08	0.11	0	0.14	0.1
ACT_COMP_SCR	0.18	0.08	0.15	0.07	0.14	0.21*	0.18	0.1	0.15	0.06	0.13	0.2*
ACT_ENGL_SCR	0.14	-0.01	0.13	0	0.09	0.18	0.14	0	0.15	-0.01	0.07	0.17
ACT_ENGWR_SCR	0.18	-0.01	0.18	0.05	0.16	0.23*	0.18	0.02	0.2	0.05	0.12	0.21
ACT_MATH_SCR	0.27**	0.08	0.19	0.15	0.3**	0.22*	0.27**	0.09	0.21*	0.15	0.29**	0.21*
ACT_READ_SCR	0.1	0.09	0.11	0.04	0	0.1	0.1	0.08	0.1	0.03	0	0.13
ACT_SCIRE_SCR	0.15	0.05	0.09	0.09	0.12	0.13	0.15	0.07	0.11	0.07	0.1	0.12
CUM_GPA	0.09	0	0.07	0	0.15	0.15	0.09	-0.01	0.06	0	0.17*	0.14
TOT_ACAD_HOURS	-0.06	-0.04	0.03	-0.04	-0.09	0	-0.06	-0.03	0.04	-0.09	-0.04	0
HS_RANK_PCT	0.16	-0.18	0.1	0.19	0.23*	0.21	0.16	-0.17	0.09	0.23*	0.22	0.18
Gender	0.09	0.09	0.06	-0.01	0.13	0.1	0.09	0.1	0.09	-0.01	0.11	0.1

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	FisherCstat	FisherCstat1	FisherCstat2	FisherCstat3	FisherCstat4	FisherCstat5	Fisherrz	Fisherrz1	Fisherrz2	Fisherrz3	Fisherrz4	Fisherrz5
extroGoldberg	0.07	-0.07	-0.02	-0.02	0.13	-0.08	0.07	-0.07	-0.03	-0.03	0.19*	-0.09
neurotGoldberg	0.02	-0.04	-0.02	-0.02	-0.01	0.01	0.02	-0.04	-0.01	-0.04	0.03	0.01
intelGoldberg	0.1	-0.07	0.07	0.04	0.16	0.08	0.1	-0.06	0.04	0.01	0.19*	0.07
agreeGoldberg	0.02	-0.04	-0.03	0.03	-0.04	-0.03	0.02	-0.05	-0.07	0.04	-0.04	-0.01
conGoldberg	0.14	0.01	0.01	0.13	0.1	0.11	0.14	0	-0.02	0.11	0.11	0.12
pvIPC7	0.15	0	0.08	-0.12	0.15	0.07	0.15	0	0.08	-0.13	0.18*	0.04
nvIPC7	-0.07	0.1	-0.06	-0.03	-0.03	-0.04	-0.07	0.1	-0.06	-0.02	-0.03	-0.03
pemIPC7	0.02	-0.06	0	-0.07	0.07	-0.09	0.02	-0.06	-0.02	-0.08	0.14	-0.1
nemIPC7	0.02	-0.05	-0.04	0.04	0	0	0.02	-0.04	-0.03	0.03	0.01	-0.01
conIPC7	0.16	0.13	0.1	0.13	0.02	0.06	0.16	0.13	0.1	0.12	0.01	0.05
agreeIPC7	0	-0.01	0.02	0.07	-0.15	-0.01	0	-0.02	0	0.09	-0.18*	0.02
cnvIPC7	-0.01	0.03	-0.01	0.05	-0.2*	-0.01	-0.01	0	0.01	0.06	-0.21*	-0.02
ReaRIASEC	0.02	0.17*	0	0.01	0.02	-0.01	0.02	0.17	0.02	-0.01	0	-0.02
InvRIASEC	-0.09	0	0	-0.05	0.03	-0.04	-0.09	0.01	0.02	-0.07	0.05	-0.08
ArtRIASEC	-0.07	-0.04	-0.03	-0.14	-0.01	-0.09	-0.07	-0.03	-0.05	-0.17	0.04	-0.1
SocRIASEC	-0.15	-0.09	0.06	-0.15	-0.06	-0.13	-0.15	-0.09	0.04	-0.18*	-0.01	-0.14
EntRIASEC	-0.03	-0.01	-0.03	-0.07	0.06	-0.06	-0.03	-0.01	-0.05	-0.07	0.07	-0.06
ConRIASEC	-0.01	0.1	-0.04	0.06	-0.02	-0.01	-0.01	0.09	-0.04	0.06	-0.04	-0.02

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	FisherCstat	FisherCstat1	FisherCstat2	FisherCstat3	FisherCstat4	FisherCstat5	Fisherrz	Fisherrz1	Fisherrz2	Fisherrz3	Fisherrz4	Fisherrz5
numpredout	-0.04	-0.17*	0.06	0.01	0.03	-0.04	-0.04	-0.18*	0.05	0.04	0.03	-0.07
numratees	-0.05	-0.12	0.11	0	-0.14	-0.03	-0.05	-0.11	0.12	0	-0.16	-0.03
lengthtimepred	-0.11	-0.08	0	0.07	-0.21*	-0.09	-0.1	-0.07	0	0.03	-0.2*	-0.1
numpredoutformal	-0.09	0.03	0.01	-0.01	-0.12	-0.11	-0.09	0.04	0	-0.02	-0.11	-0.11
numpredouttraining	0.02	0	0.11	0.01	-0.05	-0.08	0.02	0.02	0.06	0.01	-0.02	-0.08
numpredouttrainingformal	-0.07	0.02	-0.03	-0.11	-0.1	-0.17*	-0.07	0.04	-0.01	-0.13	-0.09	-0.17*
lengthtraining	-0.14	-0.15	0.03	-0.04	-0.11	-0.05	-0.14	-0.16	0	-0.03	-0.1	-0.04
amtstats	-0.01	-0.13	0.05	-0.03	0.01	-0.06	-0.01	-0.13	0.03	-0.02	0.01	-0.04
amtdecisionmkg	0.16	0.05	0.05	0.16	0.09	0.16	0.16	0.07	0.02	0.18*	0.07	0.13
TrPerExprAware	0.29**	-0.05	0.05	0.16	0.32**	0.33**	0.29**	-0.04	0.07	0.14	0.31**	0.31**
TrPerSaidUsed	0.33**	-0.06	0.12	0.22**	0.31**	0.37**	0.33**	-0.04	0.14	0.21*	0.29**	0.34**
TrPerUsedCorrectly	0.36**	-0.12	0.31**	0.39**	0.34**	0.18	0.36**	-0.11	0.32**	0.37**	0.3**	0.18
TrPerSaidUsed3rdVar	-0.05	-0.13	0.06	-0.07	0.03	0.01	-0.05	-0.14	0.04	-0.06	0.03	0.01

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	Cwtx1x2_s	Cwtx1x2_1_s	Cwtx1x2_2_s	Cwtx1x2_3_s	Cwtx1x2_4_s	Cwtx1x2_5_s	ConfAbsOverallPre	ConfAbs1	ConfAbs2	ConfAbs3	ConfAbs4	ConfAbs5	ConfAbsOverallPost
Fisherra	0.65**	0.2*	0.41**	0.54**	0.55**	0.57**	0.04	0.03	0.06	0.04	0.16	0.18*	0.15
Fisherra1	0.09	0.25**	-0.06	0.04	0.12	0.09	0.03	0.13	0.15	0.04	0.13	-0.01	0.12
Fisherra2	0.29**	0.09	0.42**	0.13	0.23**	0.22*	0.15	0.07	0.03	0.06	0.09	0.13	0.07
Fisherra3	0.24**	0.05	0.18*	0.47**	0.15	0.12	-0.19*	-0.01	-0.08	-0.01	-0.01	-0.04	-0.03
Fisherra4	0.4**	0.03	0.16	0.37**	0.47**	0.28**	-0.07	-0.18*	-0.11	-0.1	0.05	0.05	-0.02
Fisherra5	0.21*	0.02	0.23**	0.2*	0.25**	0.55**	0.07	0.02	0.06	0.08	0.02	0.18*	0.11
SS	0.4**	0.01	0.33**	0.43**	0.37**	0.38**	0.08	-0.04	0.08	0.05	0.13	0.12	0.01
SS1	0.14	0.15	0.06	0.1	0.07	0.09	-0.03	0.09	0.06	0.03	0.08	0.1	0.01
SS2	0.26**	0.02	0.29**	0.19*	0.18*	0.13	0.22*	-0.01	-0.09	-0.01	0.05	0.04	-0.04
SS3	0.16	0.03	0.26**	0.4**	0.16	0.19*	-0.01	-0.02	0.03	0.08	0.06	0.02	-0.02
SS4	0.27**	-0.08	0.15	0.3**	0.35**	0.22**	-0.01	-0.12	-0.06	-0.09	0.06	-0.01	-0.12
SS5	0.18*	-0.09	0.1	0.13	0.17*	0.2*	0.04	-0.05	0.14	0.07	0.07	0.16	0.08
FisherRslin	-0.12	0.03	-0.21*	-0.25**	-0.28**	-0.28**	0.04	0.01	-0.14	-0.04	-0.14	-0.11	-0.1
FisherRslin1	0.25**	0.13	0.1	0.12	0.23**	0.25**	0.1	0.01	-0.08	-0.12	-0.05	-0.01	-0.07
FisherRslin2	-0.12	0.04	-0.26**	-0.09	0.01	0.07	0.06	0	-0.08	-0.07	-0.03	-0.02	-0.14
FisherRslin3	-0.12	0.02	-0.12	-0.26**	-0.17*	-0.13	0.1	0.04	-0.11	0.09	0.01	0.01	-0.02
FisherRslin4	-0.2*	0.02	-0.15	-0.2*	-0.29**	-0.27**	0.01	-0.02	-0.13	-0.02	-0.02	-0.06	-0.02
FisherRslin5	-0.25**	0.09	-0.13	-0.28**	-0.31**	-0.39**	0.06	0.03	-0.09	-0.02	-0.19*	-0.06	-0.03

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	Cwtx1x2_s	Cwtx1x2_1_s	Cwtx1x2_2_s	Cwtx1x2_3_s	Cwtx1x2_4_s	Cwtx1x2_5_s	ConfAbsOverallPre	ConfAbs1	ConfAbs2	ConfAbs3	ConfAbs4	ConfAbs5	ConfAbsOverallPost
FisherG	-0.11	-0.12	0.07	0.01	0.02	0.1	0.21*	0.11	0.05	0.07	0.01	-0.01	0.06
FisherG1	-0.04	0.01	-0.12	0.03	0.12	0.04	0.1	0.13	0.08	0.09	0.17*	0.02	0.07
FisherG2	0.1	-0.05	-0.08	-0.11	0.05	0.03	0.07	0.05	0	0.06	0.07	0.09	0
FisherG3	0.01	0.05	-0.1	-0.03	-0.1	-0.01	-0.16	-0.06	-0.07	-0.01	-0.06	-0.07	0
FisherG4	0.29**	-0.05	-0.09	0.16	-0.09	-0.03	-0.17	-0.26**	-0.34**	-0.21*	-0.08	-0.16	-0.16
FisherG5	0.08	-0.03	0.03	-0.05	-0.07	0.14	-0.02	-0.02	-0.1	-0.04	-0.14	-0.04	0
FisherCstat	0.7**	0.17*	0.41**	0.53**	0.54**	0.54**	0.06	0.05	0.11	0.06	0.18*	0.18*	0.15
FisherCstat1	0.16	0.26**	0.11	-0.02	0.01	0.01	-0.14	0.07	0.1	0.09	0.01	-0.02	0.07
FisherCstat2	0.47**	0.08	0.52**	0.36**	0.23**	0.23**	0.13	0.06	0.03	0.05	0.09	0.12	0.13
FisherCstat3	0.48**	0.06	0.27**	0.59**	0.37**	0.22**	-0.12	0.02	-0.03	0	0.08	0.04	0.05
FisherCstat4	0.48**	0.03	0.27**	0.41**	0.55**	0.44**	0.11	-0.01	0.1	0.07	0.17	0.17*	0.16
FisherCstat5	0.41**	0.14	0.24**	0.3**	0.39**	0.67**	0.08	0.06	0.11	0.02	0.14	0.21*	0.09
Fisherrz	0.7**	0.17*	0.4**	0.53**	0.54**	0.54**	0.06	0.05	0.11	0.06	0.18*	0.18*	0.15
Fisherrz1	0.18*	0.23**	0.13	-0.01	0.02	0	-0.14	0.07	0.1	0.1	0.01	-0.01	0.07
Fisherrz2	0.48**	0.06	0.48**	0.38**	0.22**	0.2*	0.12	0.04	0.01	0.05	0.08	0.11	0.13
Fisherrz3	0.48**	0.03	0.25**	0.57**	0.41**	0.24**	-0.11	0	-0.04	-0.03	0.09	0.04	0.04
Fisherrz4	0.44**	0.06	0.3**	0.37**	0.5**	0.45**	0.14	0.03	0.14	0.09	0.15	0.18*	0.19*
Fisherrz5	0.42**	0.14	0.21*	0.31**	0.4**	0.66**	0.08	0.05	0.11	0.01	0.16	0.19*	0.06

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	Cwtx1x2_s	Cwtx1x2_1_s	Cwtx1x2_2_s	Cwtx1x2_3_s	Cwtx1x2_4_s	Cwtx1x2_5_s	ConfAbsOverallPre	ConfAbs1	ConfAbs2	ConfAbs3	ConfAbs4	ConfAbs5	ConfAbsOverallPost
Cwtx1x2_s	1**	0.18*	0.29**	0.43**	0.47**	0.43**	0.06	0.06	0.05	-0.01	0.14	0.11	0.04
Cwtx1x2_1_s	0.18*	1**	0.31**	-0.05	0.01	0.1	0.04	0.04	0.08	0.05	0.01	0.13	0.13
Cwtx1x2_2_s	0.29**	0.31**	1**	0.29**	0.24**	0.33**	0.13	0.17*	0.2*	0.18*	0.19*	0.25**	0.23**
Cwtx1x2_3_s	0.43**	-0.05	0.29**	1**	0.47**	0.34**	0.02	0	0.01	0.03	0.17	0.06	-0.03
Cwtx1x2_4_s	0.47**	0.01	0.24**	0.47**	1**	0.56**	0.04	0.02	0.04	-0.02	0.16	0.1	-0.07
Cwtx1x2_5_s	0.43**	0.1	0.33**	0.34**	0.56**	1**	0.16	0.18*	0.21*	0.14	0.27**	0.33**	0.12
ConfAbsOverallPre	0.06	0.04	0.13	0.02	0.04	0.16	1**	0.39**	0.39**	0.4**	0.35**	0.46**	0.41**
ConfAbs1	0.06	0.04	0.17*	0	0.02	0.18*	0.39**	1**	0.62**	0.54**	0.43**	0.49**	0.47**
ConfAbs2	0.05	0.08	0.2*	0.01	0.04	0.21*	0.39**	0.62**	1**	0.62**	0.5**	0.6**	0.59**
ConfAbs3	-0.01	0.05	0.18*	0.03	-0.02	0.14	0.4**	0.54**	0.62**	1**	0.6**	0.69**	0.69**
ConfAbs4	0.14	0.01	0.19*	0.17	0.16	0.27**	0.35**	0.43**	0.5**	0.6**	1**	0.69**	0.57**
ConfAbs5	0.11	0.13	0.25**	0.06	0.1	0.33**	0.46**	0.49**	0.6**	0.69**	0.69**	1**	0.68**
ConfAbsOverallPost	0.04	0.13	0.23**	-0.03	-0.07	0.12	0.41**	0.47**	0.59**	0.69**	0.57**	0.68**	1**
ConfRelOverallPre	0.11	0.07	0.16	0.14	0.13	0.14	0.67**	0.33**	0.4**	0.39**	0.37**	0.46**	0.33**
ConfRel1	-0.06	0.1	0.2*	0.06	0.02	0.09	0.4**	0.64**	0.55**	0.55**	0.46**	0.51**	0.41**
ConfRel2	0.05	0.11	0.25**	0.06	0.07	0.15	0.37**	0.47**	0.71**	0.6**	0.48**	0.56**	0.56**
ConfRel3	0.03	0.07	0.23**	0.08	0.02	0.12	0.35**	0.46**	0.51**	0.82**	0.54**	0.62**	0.57**
ConfRel4	0.15	0.1	0.23**	0.22**	0.15	0.25**	0.36**	0.37**	0.45**	0.58**	0.77**	0.64**	0.54**
ConfRel5	0.08	0.2*	0.26**	0.07	0.05	0.28**	0.44**	0.46**	0.58**	0.67**	0.63**	0.85**	0.62**
ConfRelOverallPost	0.04	0.22*	0.26**	0.04	0.01	0.13	0.45**	0.45**	0.56**	0.66**	0.55**	0.64**	0.76**

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	Cwtx1x2_s	Cwtx1x2_1_s	Cwtx1x2_2_s	Cwtx1x2_3_s	Cwtx1x2_4_s	Cwtx1x2_5_s	ConfAbsOverallPre	ConfAbs1	ConfAbs2	ConfAbs3	ConfAbs4	ConfAbs5	ConfAbsOverallPost
ACT_COMP_SCR	0.08	0.12	0.15	0.3**	0.25*	0.25*	0.13	0.08	0.14	0.1	0.09	0.16	-0.09
ACT_ENGL_SCR	0.1	0.09	0.09	0.18	0.17	0.17	0.13	0.07	0.15	0.11	0.09	0.23*	0
ACT_ENGWR_SCR	0.16	0.04	0.12	0.19	0.16	0.11	0.05	0.01	0.16	0.12	0.07	0.19	0
ACT_MATH_SCR	0.15	0.13	0.28**	0.32**	0.41**	0.35**	0.2	0.24*	0.21*	0.18	0.2	0.21*	-0.02
ACT_READ_SCR	0.04	0.14	0.04	0.18	0.05	0.11	0.06	-0.04	0.12	0.03	0.03	0.09	-0.09
ACT_SCIRE_SCR	0.1	0.18	0.12	0.24*	0.16	0.12	0.06	0.07	0.08	0.05	0.01	0.13	-0.16
CUM_GPA	-0.02	0.15	0.24**	0.12	0.15	0.13	0.05	-0.02	-0.05	0.02	0.12	0.07	-0.06
TOT_ACAD_HOURS	-0.01	-0.11	0.03	0.01	0	-0.05	0.09	0.1	-0.02	0.03	-0.04	0.04	0.02
HS_RANK_PCT	0.12	-0.04	0.15	0.22*	0.26*	0.25*	0.18	-0.01	-0.04	0.01	0.17	0.06	-0.04
Gender	0.01	0.03	0.08	-0.01	0.15	0.21*	0.23*	0.23**	0.18*	0.13	0.12	0.23**	0.08
extroGoldberg	-0.02	0.12	0.08	0.02	-0.08	0.05	0.2*	0.11	0.08	0.12	0.08	0.1	0.14
neurotGoldberg	-0.01	0.04	-0.17	-0.08	-0.04	-0.07	-0.11	-0.13	-0.1	-0.04	-0.14	-0.09	-0.08
intelGoldberg	-0.01	-0.01	0.16	0.18*	0.09	0.09	0.19*	0.05	0.11	0.09	0.06	0.19*	0.11
agreeGoldberg	0.02	0.01	0.04	0.15	0.05	0.11	0.14	0.05	0.12	0.06	0.1	0.09	0.04
conGoldberg	0.15	0.09	0.09	0.31**	0.13	0.11	0.12	0.08	0.05	0.1	0.13	0.11	0.11
pvIPC7	0.11	0.23**	0.2*	0.08	0.04	0.13	0.34**	0.17*	0.13	0.16	0.13	0.13	0.23**
nvIPC7	-0.17	0	-0.01	-0.26**	-0.17*	-0.14	-0.21*	0.06	0.09	-0.15	-0.08	-0.11	-0.07
pemIPC7	0.01	0.08	0.07	0.05	-0.08	0.11	0.2*	0.04	0.03	0.1	0.01	0.02	0.04
nemIPC7	-0.06	-0.01	-0.17*	0.01	0.02	-0.03	-0.21*	-0.14	-0.05	-0.07	-0.09	-0.02	-0.12
conIPC7	0.16	0.07	0.07	0.2*	0.1	0.06	-0.05	0.02	-0.04	0.04	0.06	0.02	0.07
agreeIPC7	-0.01	-0.03	-0.1	0.08	0.12	0.05	-0.18	-0.14	-0.1	-0.09	-0.11	-0.1	-0.13
cnvIPC7	0.08	0.06	0.02	0.06	-0.03	0	-0.02	0.08	0.01	0.04	0.09	0.05	-0.02

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	Cwtx1x2_s	Cwtx1x2_1_s	Cwtx1x2_2_s	Cwtx1x2_3_s	Cwtx1x2_4_s	Cwtx1x2_5_s	ConfAbsOverallPre	ConfAbs1	ConfAbs2	ConfAbs3	ConfAbs4	ConfAbs5	ConfAbsOverallPost
ReaRIASEC	0.11	0.02	0.13	-0.04	-0.05	-0.06	0.04	0.09	0.16	0.08	-0.05	0.07	0.13
InvRIASEC	0.06	0.01	0.03	-0.04	-0.05	-0.06	0.08	0.03	0.07	0.04	-0.12	-0.01	0.07
ArtRIASEC	-0.05	-0.08	0.07	0.02	-0.06	-0.05	0.05	-0.07	0.03	-0.02	-0.02	0.11	0.03
SocRIASEC	-0.13	-0.02	0.17*	-0.04	-0.18*	-0.13	0.08	-0.04	0.09	0	0	-0.02	0
EntRIASEC	-0.12	0.09	0.06	-0.05	-0.1	-0.04	0.16	-0.02	0.08	0.01	0.15	0.07	0.06
ConRIASEC	0.05	0	0	0.03	0.03	-0.06	0.01	0.15	0.11	-0.02	0.04	0.09	0.08
numpredout	-0.02	-0.27**	0.07	0.11	0.08	0	0.2*	0.18*	0.15	0.1	0.11	0.13	0.16
numratees	0.05	-0.13	-0.01	-0.01	-0.05	-0.06	0.04	0.07	0.05	0.15	0.07	0.08	0.1
lengthtimepred	-0.03	-0.09	-0.02	-0.04	-0.11	-0.1	0.06	0.11	0.11	0.13	0.08	0.16	0.15
numpredoutformal	-0.07	-0.28**	-0.12	-0.1	-0.09	-0.15	0.05	0.06	0.04	0.03	-0.04	0.02	0.04
numpredouttraining	-0.04	-0.12	0.01	0.02	0.03	-0.04	0.07	0.06	-0.02	0.02	-0.13	0.04	0.05
numpredouttrainingformal	-0.14	-0.23**	-0.1	-0.09	-0.07	-0.16	-0.04	-0.01	0.01	-0.02	-0.09	-0.02	-0.03
lengthtraining	-0.04	0.03	0.04	-0.04	-0.08	-0.05	0	-0.07	-0.19*	0.13	0.11	0.03	0.06
amtstats	-0.05	-0.14	0.02	0.05	0.1	-0.01	0.04	0.02	-0.01	-0.02	-0.07	0.01	-0.07
amtdecisionmkg	0.11	0.09	0.05	0.07	0.15	0.13	0	0.05	0.05	0.01	0.13	0.14	0.05
TrPerExprAware	0.23**	0.01	0.19*	0.29**	0.19*	0.29**	-0.01	-0.06	0.07	0.01	0.05	0.07	0.03
TrPerSaidUsed	0.31**	0	0.22*	0.31**	0.2*	0.31**	0	-0.02	0.12	0.04	0.07	0.1	0.06
TrPerUsedCorrectly	0.31**	0.05	0.24*	0.24**	0.2*	0.19*	-0.07	0.04	0.2*	0.15	0.11	0.09	0.17
TrPerSaidUsed3rdVar	0.01	0	0.02	-0.07	-0.12	0.08	0.08	0.01	0.01	0.1	0.06	0.07	0.08

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	ConfAbsOverallPost	ConfRelOverallPre	ConfRel1	ConfRel2	ConfRel3	ConfRel4	ConfRel5	ConfRelOverallPost
Fisherra	0.15	0.08	-0.01	0.07	0.09	0.17*	0.18*	0.19*
Fisherra1	0.12	-0.05	0.18*	0.1	-0.01	0.13	0.12	0.2*
Fisherra2	0.07	0.11	0.01	0.05	0.14	0.15	0.13	0.11
Fisherra3	-0.03	-0.05	-0.01	-0.04	0.01	-0.04	-0.05	-0.1
Fisherra4	-0.02	0.02	-0.18*	-0.12	-0.03	0.03	-0.04	0.01
Fisherra5	0.11	0.09	-0.01	0.06	0.04	0.04	0.13	0.11
SS	0.01	0.07	-0.02	0.09	0.09	0.13	0.13	0.09
SS1	0.01	-0.13	0.1	0.04	-0.06	0.08	0.13	-0.01
SS2	-0.04	0.12	-0.06	0.02	0.07	0.09	0.05	0.04
SS3	-0.02	0.14	0.04	0.08	0.1	0.05	0.04	0.01
SS4	-0.12	0.04	-0.16	-0.09	0.01	0.02	-0.05	-0.01
SS5	0.08	0.01	0.02	0.09	0.06	0.06	0.15	0.09
FisherRslin	-0.1	-0.04	0	-0.12	-0.01	-0.11	-0.13	-0.07
FisherRslin1	-0.07	0.08	-0.03	-0.08	-0.05	-0.08	-0.05	-0.05
FisherRslin2	-0.14	-0.08	-0.07	-0.17*	-0.07	-0.08	-0.05	-0.1
FisherRslin3	-0.02	-0.04	-0.01	-0.11	0.05	-0.01	-0.02	0.02
FisherRslin4	-0.02	-0.02	0	-0.08	0.01	-0.04	-0.05	0.01
FisherRslin5	-0.03	0	0.06	0	0	-0.09	-0.09	-0.07

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	ConfAbsOverallPost	ConfRelOverallPre	ConfRel1	ConfRel2	ConfRel3	ConfRel4	ConfRel5	ConfRelOverallPost
FisherG	0.06	0.1	0.1	0.01	0.03	-0.05	-0.02	0.03
FisherG1	0.07	-0.02	0.15	0.04	0.04	0.11	0.09	0.12
FisherG2	0	0.04	0	-0.06	0.09	0.08	0.09	0.1
FisherG3	0	-0.11	-0.09	-0.06	-0.03	-0.14	-0.1	-0.13
FisherG4	-0.16	-0.16	-0.27**	-0.23**	-0.11	-0.07	-0.16	-0.16
FisherG5	0	0.03	-0.08	-0.04	-0.08	-0.15	-0.1	-0.05
FisherCstat	0.15	0.11	-0.01	0.1	0.1	0.17	0.19*	0.17*
FisherCstat1	0.07	-0.12	0.07	0.09	0.04	0	0.04	0.09
FisherCstat2	0.13	0.14	0.02	0.13	0.12	0.15	0.13	0.11
FisherCstat3	0.05	0.03	-0.11	-0.05	0.03	0.09	0.03	0
FisherCstat4	0.16	0.16	-0.02	0.05	0.06	0.13	0.09	0.12
FisherCstat5	0.09	0.05	0.02	0.05	0	0.09	0.18*	0.11
Fisherrz	0.15	0.11	-0.01	0.1	0.1	0.16	0.19*	0.17
Fisherrz1	0.07	-0.11	0.05	0.11	0.06	0	0.04	0.08
Fisherrz2	0.13	0.14	0.01	0.12	0.12	0.14	0.12	0.11
Fisherrz3	0.04	0.02	-0.12	-0.05	0.02	0.1	0.02	0
Fisherrz4	0.19*	0.16	0.03	0.09	0.08	0.14	0.12	0.14
Fisherrz5	0.06	0.03	0.01	0.05	0	0.09	0.17*	0.1

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	ConfAbsOverallPost	ConfRelOverallPre	ConfRel1	ConfRel2	ConfRel3	ConfRel4	ConfRel5	ConfRelOverallPost
Cwtx1x2_s	0.04	0.11	-0.06	0.05	0.03	0.15	0.08	0.04
Cwtx1x2_1_s	0.13	0.07	0.1	0.11	0.07	0.1	0.2*	0.22*
Cwtx1x2_2_s	0.23**	0.16	0.2*	0.25**	0.23**	0.23**	0.26**	0.26**
Cwtx1x2_3_s	-0.03	0.14	0.06	0.06	0.08	0.22**	0.07	0.04
Cwtx1x2_4_s	-0.07	0.13	0.02	0.07	0.02	0.15	0.05	0.01
Cwtx1x2_5_s	0.12	0.14	0.09	0.15	0.12	0.25**	0.28**	0.13
ConfAbsOverallPre	0.41**	0.67**	0.4**	0.37**	0.35**	0.36**	0.44**	0.45**
ConfAbs1	0.47**	0.33**	0.64**	0.47**	0.46**	0.37**	0.46**	0.45**
ConfAbs2	0.59**	0.4**	0.55**	0.71**	0.51**	0.45**	0.58**	0.56**
ConfAbs3	0.69**	0.39**	0.55**	0.6**	0.82**	0.58**	0.67**	0.66**
ConfAbs4	0.57**	0.37**	0.46**	0.48**	0.54**	0.77**	0.63**	0.55**
ConfAbs5	0.68**	0.46**	0.51**	0.56**	0.62**	0.64**	0.85**	0.64**
ConfAbsOverallPost	1**	0.33**	0.41**	0.56**	0.57**	0.54**	0.62**	0.76**
ConfRelOverallPre	0.33**	1**	0.53**	0.54**	0.5**	0.46**	0.5**	0.52**
ConfRel1	0.41**	0.53**	1**	0.69**	0.59**	0.58**	0.64**	0.61**
ConfRel2	0.56**	0.54**	0.69**	1**	0.7**	0.62**	0.68**	0.68**
ConfRel3	0.57**	0.5**	0.59**	0.7**	1**	0.68**	0.71**	0.73**
ConfRel4	0.54**	0.46**	0.58**	0.62**	0.68**	1**	0.74**	0.68**
ConfRel5	0.62**	0.5**	0.64**	0.68**	0.71**	0.74**	1**	0.77**
ConfRelOverallPost	0.76**	0.52**	0.61**	0.68**	0.73**	0.68**	0.77**	1**

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	ConfAbsOverallPost	ConfRelOverallPre	ConfRel1	ConfRel2	ConfRel3	ConfRel4	ConfRel5	ConfRelOverallPost
ACT_COMP_SCR	-0.09	0.27*	0.18	0.2*	0.22*	0.18	0.23*	0.2*
ACT_ENGL_SCR	0	0.2	0.18	0.19	0.21*	0.15	0.24*	0.22*
ACT_ENGWR_SCR	0	0.17	0.11	0.16	0.23*	0.12	0.19	0.2
ACT_MATH_SCR	-0.02	0.26*	0.34**	0.23*	0.21*	0.18	0.24*	0.19
ACT_READ_SCR	-0.09	0.2	0.09	0.2*	0.17	0.13	0.18	0.18
ACT_SCIRE_SCR	-0.16	0.19	0.16	0.12	0.14	0.11	0.17	0.08
CUM_GPA	-0.06	0.05	0.15	0.05	0.04	0.07	0.09	0.02
TOT_ACAD_HOURS	0.02	0	0.12	-0.04	0.06	-0.04	0.02	0.03
HS_RANK_PCT	-0.04	0.24*	0.09	0.06	0.06	0.07	0.06	0.08
Gender	0.08	0.27**	0.19*	0.18*	0.16	0.12	0.17	0.15
extroGoldberg	0.14	0.05	0.16	0.12	0.14	0.13	0.19*	0.2*
neurotGoldberg	-0.08	-0.12	-0.13	-0.05	-0.03	-0.16	-0.05	-0.12
intelGoldberg	0.11	0.18	0.08	0.07	0.1	0.16	0.14	0.14
agreeGoldberg	0.04	0.1	0.08	0.09	0.06	0.04	0.11	0.07
conGoldberg	0.11	0.13	0.04	0	0.11	0.18*	0.16	0.17
pvIPC7	0.23**	0.21*	0.22**	0.16	0.13	0.14	0.21*	0.34**
nvIPC7	-0.07	-0.16	-0.06	0.05	-0.03	-0.09	-0.13	-0.14
pemIPC7	0.04	0.01	0.1	0.12	0.09	0.07	0.11	0.12
nemIPC7	-0.12	-0.19*	-0.14	-0.1	-0.11	-0.16	-0.03	-0.13
conIPC7	0.07	0.02	-0.02	-0.04	0.07	0.11	0.07	0.08
agreeIPC7	-0.13	-0.12	-0.13	-0.1	-0.17	-0.18*	-0.13	-0.12
cnvIPC7	-0.02	-0.06	0.03	0	-0.04	0.05	0.09	0.03

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	ConfAbsOverallPost	ConfRelOverallPre	ConfRel1	ConfRel2	ConfRel3	ConfRel4	ConfRel5	ConfRelOverallPost
ReaRIASEC	0.13	0.08	0.06	0.08	0.04	0.06	0.09	0.07
InvRIASEC	0.07	0.07	-0.04	0.04	-0.02	-0.07	0.06	0.06
ArtRIASEC	0.03	0.05	0	0.01	0.01	0.06	0.05	-0.01
SocRIASEC	0	-0.04	-0.03	0.03	-0.04	0.02	-0.02	-0.03
EntRIASEC	0.06	0.18	0.06	0.1	0.07	0.17	0.08	0.12
ConRIASEC	0.08	-0.04	0.02	0.03	0.02	0.08	0.12	0.07
numpredout	0.16	0.22*	0.24**	0.25**	0.11	0.21*	0.11	0.09
numratees	0.1	0.01	-0.02	0.04	0.08	0.11	0.04	0.01
lengthtimepred	0.15	0.08	0.05	0.04	0.07	0.11	0.1	0.04
numpredoutformal	0.04	0.02	0	0.03	-0.01	-0.05	-0.01	-0.07
numpredouttraining	0.05	0.09	0.03	0.01	0.09	-0.03	0.06	0.14
numpredouttrainingformal	-0.03	-0.01	-0.06	0.01	-0.03	-0.1	-0.05	-0.01
lengthtraining	0.06	0.04	0.05	-0.12	0.07	0.15	-0.02	-0.01
amtstats	-0.07	0.04	0.03	0.02	0.02	-0.06	0	-0.04
amtdecisionmkg	0.05	-0.01	0.08	0.11	0.07	0.13	0.09	-0.01
TrPerExprAware	0.03	-0.01	-0.11	-0.01	-0.06	-0.02	0.05	0.03
TrPerSaidUsed	0.06	0.05	-0.08	0.07	-0.01	0.02	0.07	0.04
TrPerUsedCorrectly	0.17	-0.02	0.02	0.19*	0.1	0.05	0.11	0.15
TrPerSaidUsed3rdVar	0.08	-0.03	-0.03	-0.08	0.08	0.09	0.03	0.09

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	ACT_COMP_SCR	ACT_ENGL_SCR	ACT_ENGWR_SCR	ACT_MATH_SCR	ACT_READ_SCR	ACT_SCIRE_SCR	CUM_GPA	TOT_ACAD_HOURS	HS_RANK_PCT	Gender	extroGoldberg	neurotGoldberg	intelGoldberg	agreeGoldberg	conGoldberg
Fisherra	0.21*	0.18	0.19	0.32**	0.1	0.15	0.09	-0.08	0.18	0.11	0.09	-0.03	0.12	0.04	0.1
Fisherra1	0.11	0.04	0	0.13	0.09	0.07	0.16	-0.13	-0.11	0.06	-0.03	-0.09	-0.11	0.06	0.12
Fisherra2	0.1	0.01	-0.03	0.09	0.08	0.02	0.06	-0.02	0.16	-0.04	0.07	-0.08	0.13	0	0.04
Fisherra3	0.02	0.02	0.02	0.17	0.01	0.03	0.04	0.22*	0.07	0.04	0.03	0.06	0.13	0.07	0.1
Fisherra4	0.08	0.07	0.19	0.15	-0.03	0.08	0.02	-0.19*	0.12	0.02	0.01	-0.03	0.02	-0.04	0.05
Fisherra5	0.21*	0.24*	0.26*	0.2	0.07	0.14	0.03	-0.02	0.24*	0.14	0.07	-0.01	0.11	-0.01	-0.02
SS	0.38**	0.33**	0.3**	0.39**	0.31**	0.26*	0.1	0.04	0.26*	0.03	0.06	-0.12	0.15	0.07	0.08
SS1	-0.05	0	-0.03	0.26*	0.05	0.07	0.12	-0.13	-0.02	0.06	-0.09	-0.1	-0.08	0.05	0.02
SS2	0.21*	0.11	0	0.15	0.18	0.11	0	0.07	0.14	0.11	0.11	-0.03	0.15	-0.08	0.03
SS3	0.17	0.18	0.15	0.18	0.14	0.07	0.14	0.12	0.16	0.09	-0.02	0.02	0.06	0.02	0.12
SS4	0.25*	0.23*	0.28**	0.2*	0.18	0.14	0.09	-0.06	0.2	-0.05	0	0	0.15	0	0.04
SS5	0.24*	0.2	0.24*	0.2*	0.14	0.21*	-0.06	0.08	0.12	-0.07	0.09	-0.1	0.08	0.1	-0.01
FisherRslin	0.09	0.11	0.13	-0.12	0.19	0.1	-0.06	0.15	-0.2	-0.11	-0.13	0.03	0	-0.02	-0.12
FisherRslin1	0.37**	0.38**	0.43**	0.24*	0.33**	0.33**	0.06	-0.05	0.16	-0.07	-0.15	0.06	-0.01	0.06	-0.06
FisherRslin2	0.14	0.12	0.15	0	0.17	0.07	0.11	0.09	0.07	-0.12	-0.11	0.11	0.07	-0.01	0.04
FisherRslin3	0.07	0.12	0.12	-0.12	0.17	0.09	0.05	0.18*	-0.17	-0.17	-0.05	-0.01	0.01	-0.01	0.03
FisherRslin4	0.01	0.02	0	-0.15	0.15	0.02	-0.04	0.15	-0.16	-0.14	-0.11	0.03	0.05	0.01	-0.03
FisherRslin5	-0.23*	-0.16	-0.16	-0.28**	-0.08	-0.15	-0.12	0.11	-0.44**	0.01	0.01	0	0	-0.04	-0.14

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	ACT_COMP_SCR	ACT_ENGL_SCR	ACT_ENGWR_SCR	ACT_MATH_SCR	ACT_READ_SCR	ACT_SCIRE_SCR	CUM_GPA	TOT_ACAD_HOURS	HS_RANK_PCT	Gender	extroGoldberg	neurotGoldberg	intelGoldberg	agreeGoldberg	conGoldberg
FisherG	0.07	0.08	0.06	0.11	0.1	-0.04	0.08	0.06	0.04	-0.03	0.06	-0.03	0.12	0.17	-0.04
FisherG1	0.14	0.14	0.14	0.17	0.08	0.12	0.06	-0.08	-0.06	0.01	-0.06	-0.07	-0.11	0.07	0.03
FisherG2	0.04	-0.03	-0.06	-0.06	0.06	0	-0.09	0.01	0.11	-0.07	-0.01	0.06	-0.02	-0.1	0
FisherG3	-0.16	-0.1	-0.11	-0.04	-0.14	-0.09	-0.08	0.08	-0.12	0	0.08	0.1	0.02	0.08	-0.03
FisherG4	-0.12	-0.11	0.03	-0.12	0	-0.06	-0.08	0.06	0.05	-0.2*	0.01	0.01	-0.02	0.06	0.13
FisherG5	-0.04	0.05	0.06	-0.11	-0.03	-0.11	-0.09	-0.04	0.09	0.06	0.01	0.03	-0.01	0.01	-0.09
FisherCstat	0.18	0.14	0.18	0.27**	0.1	0.15	0.09	-0.06	0.16	0.09	0.07	0.02	0.1	0.02	0.14
FisherCstat1	0.08	-0.01	-0.01	0.08	0.09	0.05	0	-0.04	-0.18	0.09	-0.07	-0.04	-0.07	-0.04	0.01
FisherCstat2	0.15	0.13	0.18	0.19	0.11	0.09	0.07	0.03	0.1	0.06	-0.02	-0.02	0.07	-0.03	0.01
FisherCstat3	0.07	0	0.05	0.15	0.04	0.09	0	-0.04	0.19	-0.01	-0.02	-0.02	0.04	0.03	0.13
FisherCstat4	0.14	0.09	0.16	0.3**	0	0.12	0.15	-0.09	0.23*	0.13	0.13	-0.01	0.16	-0.04	0.1
FisherCstat5	0.21*	0.18	0.23*	0.22*	0.1	0.13	0.15	0	0.21	0.1	-0.08	0.01	0.08	-0.03	0.11
Fisherrz	0.18	0.14	0.18	0.27**	0.1	0.15	0.09	-0.06	0.16	0.09	0.07	0.02	0.1	0.02	0.14
Fisherrz1	0.1	0	0.02	0.09	0.08	0.07	-0.01	-0.03	-0.17	0.1	-0.07	-0.04	-0.06	-0.05	0
Fisherrz2	0.15	0.15	0.2	0.21*	0.1	0.11	0.06	0.04	0.09	0.09	-0.03	-0.01	0.04	-0.07	-0.02
Fisherrz3	0.06	-0.01	0.05	0.15	0.03	0.07	0	-0.09	0.23*	-0.01	-0.03	-0.04	0.01	0.04	0.11
Fisherrz4	0.13	0.07	0.12	0.29**	0	0.1	0.17*	-0.04	0.22	0.11	0.19*	0.03	0.19*	-0.04	0.11
Fisherrz5	0.2*	0.17	0.21	0.21*	0.13	0.12	0.14	0	0.18	0.1	-0.09	0.01	0.07	-0.01	0.12

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	ACT_COMP_SCR	ACT_ENGL_SCR	ACT_ENGWR_SCR	ACT_MATH_SCR	ACT_READ_SCR	ACT_SCIRE_SCR	CUM_GPA	TOT_ACAD_HOURS	HS_RANK_PCT	Gender	extroGoldberg	neurotGoldberg	intelGoldberg	agreeGoldberg	conGoldberg
Cwtx1x2_s	0.08	0.1	0.16	0.15	0.04	0.1	-0.02	-0.01	0.12	0.01	-0.02	-0.01	-0.01	0.02	0.15
Cwtx1x2_1_s	0.12	0.09	0.04	0.13	0.14	0.18	0.15	-0.11	-0.04	0.03	0.12	0.04	-0.01	0.01	0.09
Cwtx1x2_2_s	0.15	0.09	0.12	0.28**	0.04	0.12	0.24**	0.03	0.15	0.08	0.08	-0.17	0.16	0.04	0.09
Cwtx1x2_3_s	0.3**	0.18	0.19	0.32**	0.18	0.24*	0.12	0.01	0.22*	-0.01	0.02	-0.08	0.18*	0.15	0.31**
Cwtx1x2_4_s	0.25*	0.17	0.16	0.41**	0.05	0.16	0.15	0	0.26*	0.15	-0.08	-0.04	0.09	0.05	0.13
Cwtx1x2_5_s	0.25*	0.17	0.11	0.35**	0.11	0.12	0.13	-0.05	0.25*	0.21*	0.05	-0.07	0.09	0.11	0.11
ConfAbsOverallPre	0.13	0.13	0.05	0.2	0.06	0.06	0.05	0.09	0.18	0.23*	0.2*	-0.11	0.19*	0.14	0.12
ConfAbs1	0.08	0.07	0.01	0.24*	-0.04	0.07	-0.02	0.1	-0.01	0.23**	0.11	-0.13	0.05	0.05	0.08
ConfAbs2	0.14	0.15	0.16	0.21*	0.12	0.08	-0.05	-0.02	-0.04	0.18*	0.08	-0.1	0.11	0.12	0.05
ConfAbs3	0.1	0.11	0.12	0.18	0.03	0.05	0.02	0.03	0.01	0.13	0.12	-0.04	0.09	0.06	0.1
ConfAbs4	0.09	0.09	0.07	0.2	0.03	0.01	0.12	-0.04	0.17	0.12	0.08	-0.14	0.06	0.1	0.13
ConfAbs5	0.16	0.23*	0.19	0.21*	0.09	0.13	0.07	0.04	0.06	0.23**	0.1	-0.09	0.19*	0.09	0.11
ConfAbsOverallPost	-0.09	0	0	-0.02	-0.09	-0.16	-0.06	0.02	-0.04	0.08	0.14	-0.08	0.11	0.04	0.11
ConfRelOverallPre	0.27*	0.2	0.17	0.26*	0.2	0.19	0.05	0	0.24*	0.27**	0.05	-0.12	0.18	0.1	0.13
ConfRel1	0.18	0.18	0.11	0.34**	0.09	0.16	0.15	0.12	0.09	0.19*	0.16	-0.13	0.08	0.08	0.04
ConfRel2	0.2*	0.19	0.16	0.23*	0.2*	0.12	0.05	-0.04	0.06	0.18*	0.12	-0.05	0.07	0.09	0
ConfRel3	0.22*	0.21*	0.23*	0.21*	0.17	0.14	0.04	0.06	0.06	0.16	0.14	-0.03	0.1	0.06	0.11
ConfRel4	0.18	0.15	0.12	0.18	0.13	0.11	0.07	-0.04	0.07	0.12	0.13	-0.16	0.16	0.04	0.18*
ConfRel5	0.23*	0.24*	0.19	0.24*	0.18	0.17	0.09	0.02	0.06	0.17	0.19*	-0.05	0.14	0.11	0.16
ConfRelOverallPost	0.2*	0.22*	0.2	0.19	0.18	0.08	0.02	0.03	0.08	0.15	0.2*	-0.12	0.14	0.07	0.17

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	ACT_COMP_SCR	ACT_ENGL_SCR	ACT_ENGWR_SCR	ACT_MATH_SCR	ACT_READ_SCR	ACT_SCIRE_SCR	CUM_GPA	TOT_ACAD_HOURS	HS_RANK_PCT	Gender	extroGoldberg	neurotGoldberg	intelGoldberg	agreeGoldberg	conGoldberg
ACT_COMP_SCR	1**	0.86**	0.83**	0.63**	0.84**	0.79**	0.39**	-0.13	0.31**	0.14	-0.13	-0.06	0.15	0.04	0.08
ACT_ENGL_SCR	0.86**	1**	0.93**	0.41**	0.68**	0.63**	0.34**	-0.04	0.2	0.05	-0.16	-0.05	0.1	0.04	-0.03
ACT_ENGWR_SCR	0.83**	0.93**	1**	0.38**	0.66**	0.62**	0.33**	-0.03	0.12	0.04	-0.25*	-0.05	0.06	-0.02	0.03
ACT_MATH_SCR	0.63**	0.41**	0.38**	1**	0.4**	0.51**	0.25*	-0.05	0.55**	0.39**	0.02	-0.13	0.05	-0.03	0.01
ACT_READ_SCR	0.84**	0.68**	0.66**	0.4**	1**	0.59**	0.28**	-0.1	0.21	0.01	-0.11	0.02	0.14	0.03	0.04
ACT_SCIRE_SCR	0.79**	0.63**	0.62**	0.51**	0.59**	1**	0.35**	-0.13	0.21	0.08	-0.19	-0.01	0.06	-0.15	-0.06
CUM_GPA	0.39**	0.34**	0.33**	0.25*	0.28**	0.35**	1**	0.04	0.28*	-0.03	-0.03	-0.08	0.01	-0.01	0.18*
TOT_ACAD_HOURS	-0.13	-0.04	-0.03	-0.05	-0.1	-0.13	0.04	1**	-0.17	-0.03	0.01	0	0.25**	-0.1	0.16
HS_RANK_PCT	0.31**	0.2	0.12	0.55**	0.21	0.21	0.28*	-0.17	1**	0.04	0.04	-0.1	-0.18	-0.03	0.12
Gender	0.14	0.05	0.04	0.39**	0.01	0.08	-0.03	-0.03	0.04	1**	-0.02	-0.18*	-0.1	-0.2*	-0.05
extroGoldberg	-0.13	-0.16	-0.25*	0.02	-0.11	-0.19	-0.03	0.01	0.04	-0.02	1**	-0.16	0.23**	0.28**	0.15
neurotGoldberg	-0.06	-0.05	-0.05	-0.13	0.02	-0.01	-0.08	0	-0.1	-0.18*	-0.16	1**	0.03	-0.18*	-0.15
intelGoldberg	0.15	0.1	0.06	0.05	0.14	0.06	0.01	0.25**	-0.18	-0.1	0.23**	0.03	1**	0.3**	0.31**
agreeGoldberg	0.04	0.04	-0.02	-0.03	0.03	-0.15	-0.01	-0.1	-0.03	-0.2*	0.28**	-0.18*	0.3**	1**	0.4**
conGoldberg	0.08	-0.03	0.03	0.01	0.04	-0.06	0.18*	0.16	0.12	-0.05	0.15	-0.15	0.31**	0.4**	1**
pvIPC7	0.08	-0.02	-0.05	0.27**	-0.05	-0.03	0	-0.1	0.11	0.17	0.39**	-0.09	0.22*	0.23**	0.28**
nvIPC7	0.02	-0.01	0.03	-0.01	0.03	0.08	0	-0.08	-0.03	0.16	-0.2*	0.05	-0.13	-0.4**	-0.3**

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	ACT_COMP_SCR	ACT_ENGL_SCR	ACT_ENGWR_SCR	ACT_MATH_SCR	ACT_READ_SCR	ACT_SCIRE_SCR	CUM_GPA	TOT_ACAD_HOURS	HS_RANK_PCT	Gender	extroGoldberg	neurotGoldberg	intelGoldberg	agreeGoldberg	conGoldberg
pemIPC7	-0.15	-0.17	-0.31**	0.02	-0.18	-0.2	-0.04	0.01	-0.02	-0.05	0.82**	-0.02	0.24**	0.4**	0.15
nemIPC7	0.02	0.07	0.08	-0.09	0.06	0.08	-0.04	0.01	0.07	-0.24**	-0.18*	0.69**	0.05	-0.11	-0.09
conIPC7	-0.01	-0.15	-0.02	-0.01	-0.08	-0.03	0.22*	0.04	-0.01	-0.01	-0.19*	0.08	0	0.16	0.7**
agreeIPC7	-0.01	0.04	0.06	-0.13	-0.05	-0.1	-0.06	-0.2*	-0.08	-0.22*	-0.24**	-0.08	-0.2*	0.42**	-0.06
cnvIPC7	-0.17	-0.11	-0.15	-0.09	-0.19	-0.03	0.08	-0.12	-0.03	0.13	-0.11	-0.1	-0.39**	0.02	0.16
ReaRIASEC	0.1	-0.02	0.03	0.17	0.08	0.2	-0.15	0.13	-0.15	0.21*	-0.07	-0.14	0.13	-0.15	0.01
InvRIASEC	-0.02	-0.04	-0.03	0.07	-0.02	0.03	-0.2*	0.3**	-0.18	-0.13	0.16	-0.04	0.13	0	0.05
ArtRIASEC	0.11	0.08	0.06	0.05	0.11	0.11	-0.12	0.18*	-0.08	-0.15	0.12	0.14	0.54**	0.14	0
SocRIASEC	-0.03	-0.06	-0.13	-0.04	0	-0.03	0.05	0.18*	-0.03	-0.34**	0.3**	0.04	0.39**	0.33**	0.14
EntRIASEC	0.24*	0.09	0.09	0.19	0.18	0.17	-0.04	-0.41**	0.23	0.17*	0.21*	-0.12	0.1	0.14	0.01
ConRIASEC	0.13	-0.05	-0.03	0.27**	0	0.11	-0.05	0	0.18	0.15	-0.15	-0.11	-0.06	-0.08	0.2*

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	ACT_COMP_SCR	ACT_ENGL_SCR	ACT_ENGWR_SCR	ACT_MATH_SCR	ACT_READ_SCR	ACT_SCIRE_SCR	CUM_GPA	TOT_ACAD_HOURS	HS_RANK_PCT	Gender	extroGoldberg	neurotGoldberg	intelGoldberg	agreeGoldberg	conGoldberg
numpredout	-0.02	0.04	0.02	-0.08	-0.02	-0.17	-0.14	0.06	-0.02	-0.04	0.07	-0.05	0.13	0.1	0.08
numratees	-0.2*	-0.11	-0.03	-0.14	-0.18	-0.29**	-0.23**	0.02	-0.3**	0.01	0.01	-0.03	-0.11	0.02	-0.02
lengthtimepred	-0.05	0.04	0.17	-0.15	0.02	-0.04	-0.1	0.21*	-0.37**	-0.08	-0.07	-0.05	0.14	-0.02	0.15
numpredoutformal	-0.07	-0.12	-0.04	-0.08	-0.05	-0.15	-0.08	0.08	0.02	-0.03	-0.01	0.01	-0.08	0.06	0.02
numpredouttraining	0.02	-0.04	0.02	-0.05	-0.04	-0.06	-0.17*	0.13	-0.1	0.01	0.1	-0.07	0.13	0.16	0.06
numpredouttrainingformal	-0.14	-0.12	-0.04	-0.05	-0.17	-0.09	-0.16	0.09	-0.14	0.1	0.07	-0.03	0.02	0.08	-0.07
lengthtraining	-0.07	0.02	0	-0.07	-0.04	-0.16	-0.05	0.05	0.01	-0.09	-0.1	-0.02	0.04	-0.11	-0.07
amtstats	-0.1	-0.13	-0.13	0	-0.06	-0.13	-0.11	0.27**	0.11	-0.2*	0.04	-0.03	0.17*	0.2*	0.11
amtdecisionmkg	-0.09	-0.12	-0.1	0.03	-0.07	-0.13	0	-0.02	0.12	0.03	0.03	0.03	0.1	-0.05	0.05
TrPerExprAware	0.06	0.04	0.09	0.05	-0.01	0.01	0.03	0.07	-0.03	0.06	-0.02	-0.03	0.05	0.04	0.15
TrPerSaidUsed	0.11	0.07	0.1	0.13	0.07	0.03	0.01	0.06	0.06	0.04	-0.06	-0.05	0.06	0.06	0.07
TrPerUsedCorrectly	0.04	0.07	0.15	0.05	0.02	0	-0.01	-0.12	-0.04	0.09	-0.03	0.15	-0.1	0.04	-0.02
TrPerSaidUsed3rdVar	0	0.03	0.04	0.01	0.03	-0.06	0	-0.05	0.13	0.04	-0.04	0.01	0.06	-0.1	0.05

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	pvIPC7	nvIPC7	pemIPC7	nemIPC7	conIPC7	agreeIPC7	cnvIPC7	ReaRIASEC	InvRIASEC	ArtRIASEC	SocRIASEC	EntRIASEC	ConRIASEC
Fisherra	0.15	-0.1	0.04	-0.02	0.11	0.05	-0.01	-0.04	-0.09	-0.03	-0.17	-0.03	-0.09
Fisherra1	0.14	-0.06	-0.04	-0.13	0.17	0.05	0.15	0.06	-0.05	-0.14	-0.09	0.02	0.06
Fisherra2	0.12	-0.01	0.09	-0.1	-0.02	0.03	-0.06	-0.05	-0.06	0.06	0.09	0.11	0.03
Fisherra3	-0.14	0	0.02	0.13	0.08	0.12	-0.06	-0.01	-0.07	0.02	-0.01	-0.15	-0.05
Fisherra4	0.02	-0.02	-0.05	-0.02	0.06	0	-0.15	-0.02	0	-0.03	-0.18*	0.01	-0.04
Fisherra5	0.17	-0.13	0.06	0.01	0	-0.06	0.05	-0.06	0.06	0.01	-0.06	-0.02	-0.12
SS	0.05	-0.13	0.05	0	0.01	0.01	-0.09	0.03	-0.01	0.09	0.02	-0.04	-0.14
SS1	0.1	-0.12	0	-0.08	0.11	0.07	0.2*	0.05	0	-0.04	0	0.04	-0.07
SS2	0.03	0.02	0.09	-0.1	-0.05	-0.02	-0.04	0	-0.05	0.08	0.02	0.01	-0.07
SS3	-0.13	-0.06	-0.04	0.02	0.09	0.04	-0.07	0.03	-0.09	0.06	0.04	-0.06	-0.11
SS4	0.06	0.01	-0.05	0.05	-0.04	-0.06	-0.17	-0.06	-0.03	0.02	-0.03	0.09	0
SS5	0.02	-0.14	0.07	0.07	-0.04	0.01	-0.09	0.03	0.06	0.09	0.01	-0.12	-0.11
FisherRslin	-0.11	-0.04	-0.03	-0.1	-0.09	0.13	0.02	0.08	0.14	0.12	0.07	-0.05	-0.08
FisherRslin1	-0.01	-0.09	-0.13	0.07	0.01	0.14	-0.01	-0.1	-0.06	0	-0.04	-0.03	-0.07
FisherRslin2	-0.09	-0.07	-0.11	0.01	0.07	0.02	-0.11	-0.15	-0.05	0	-0.09	-0.11	-0.09
FisherRslin3	-0.05	-0.15	0	-0.08	0	0.04	0.06	-0.03	0.07	0.04	0.15	-0.07	-0.13
FisherRslin4	-0.14	-0.06	-0.07	-0.12	-0.04	0.17*	0.05	0.02	0.08	0.06	0.12	-0.05	-0.07
FisherRslin5	-0.07	0.08	0.05	-0.15	-0.13	0.04	0.04	0.14	0.13	0.02	0.05	-0.04	-0.07
FisherG	0.02	-0.02	0.05	-0.09	-0.05	0.08	-0.13	-0.16	0.1	0	0.04	-0.12	-0.22*
FisherG1	0.19*	-0.07	-0.05	-0.09	0.05	0.02	-0.01	0.02	-0.02	-0.05	-0.09	0.01	0
FisherG2	0.03	0.01	-0.04	0	0.03	-0.09	-0.04	-0.11	0	0.04	-0.02	0.03	0.07
FisherG3	-0.02	0.01	0.03	0.03	-0.04	0.1	-0.01	-0.02	0.04	0.12	0.13	-0.03	-0.12
FisherG4	-0.08	0.01	0.04	-0.04	0.1	0.03	-0.16	0.02	0	0.06	-0.12	-0.07	-0.03
FisherG5	0.14	-0.08	0.05	-0.03	-0.06	0.05	0.14	-0.08	0.1	-0.05	-0.07	-0.07	-0.16

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	pvIPC7	nvIPC7	pemIPC7	nemIPC7	conIPC7	agreeIPC7	cnvIPC7	ReaRIASEC	InvRIASEC	ArtRIASEC	SocRIASEC	EntRIASEC	ConRIASEC
FisherCstat	0.15	-0.07	0.02	0.02	0.16	0	-0.01	0.02	-0.09	-0.07	-0.15	-0.03	-0.01
FisherCstat1	0	0.1	-0.06	-0.05	0.13	-0.01	0.03	0.17*	0	-0.04	-0.09	-0.01	0.1
FisherCstat2	0.08	-0.06	0	-0.04	0.1	0.02	-0.01	0	0	-0.03	0.06	-0.03	-0.04
FisherCstat3	-0.12	-0.03	-0.07	0.04	0.13	0.07	0.05	0.01	-0.05	-0.14	-0.15	-0.07	0.06
FisherCstat4	0.15	-0.03	0.07	0	0.02	-0.15	-0.2*	0.02	0.03	-0.01	-0.06	0.06	-0.02
FisherCstat5	0.07	-0.04	-0.09	0	0.06	-0.01	-0.01	-0.01	-0.04	-0.09	-0.13	-0.06	-0.01
Fisherrz	0.15	-0.07	0.02	0.02	0.16	0	-0.01	0.02	-0.09	-0.07	-0.15	-0.03	-0.01
Fisherrz1	0	0.1	-0.06	-0.04	0.13	-0.02	0	0.17	0.01	-0.03	-0.09	-0.01	0.09
Fisherrz2	0.08	-0.06	-0.02	-0.03	0.1	0	0.01	0.02	0.02	-0.05	0.04	-0.05	-0.04
Fisherrz3	-0.13	-0.02	-0.08	0.03	0.12	0.09	0.06	-0.01	-0.07	-0.17	-0.18*	-0.07	0.06
Fisherrz4	0.18*	-0.03	0.14	0.01	0.01	-0.18*	-0.21*	0	0.05	0.04	-0.01	0.07	-0.04
Fisherrz5	0.04	-0.03	-0.1	-0.01	0.05	0.02	-0.02	-0.02	-0.08	-0.1	-0.14	-0.06	-0.02
Cwtx1x2_s	0.11	-0.17	0.01	-0.06	0.16	-0.01	0.08	0.11	0.06	-0.05	-0.13	-0.12	0.05
Cwtx1x2_1_s	0.23**	0	0.08	-0.01	0.07	-0.03	0.06	0.02	0.01	-0.08	-0.02	0.09	0
Cwtx1x2_2_s	0.2*	-0.01	0.07	-0.17*	0.07	-0.1	0.02	0.13	0.03	0.07	0.17*	0.06	0
Cwtx1x2_3_s	0.08	-0.26**	0.05	0.01	0.2*	0.08	0.06	-0.04	-0.04	0.02	-0.04	-0.05	0.03
Cwtx1x2_4_s	0.04	-0.17*	-0.08	0.02	0.1	0.12	-0.03	-0.05	-0.05	-0.06	-0.18*	-0.1	0.03
Cwtx1x2_5_s	0.13	-0.14	0.11	-0.03	0.06	0.05	0	-0.06	-0.06	-0.05	-0.13	-0.04	-0.06

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	pvIPC7	nvIPC7	pemIPC7	nemIPC7	conIPC7	agreeIPC7	cnvIPC7	ReaRIASEC	InvRIASEC	ArtRIASEC	SocRIASEC	EntRIASEC	ConRIASEC
ConfAbsOverallPre	0.34**	-0.21*	0.2*	-0.21*	-0.05	-0.18	-0.02	0.04	0.08	0.05	0.08	0.16	0.01
ConfAbs1	0.17*	0.06	0.04	-0.14	0.02	-0.14	0.08	0.09	0.03	-0.07	-0.04	-0.02	0.15
ConfAbs2	0.13	0.09	0.03	-0.05	-0.04	-0.1	0.01	0.16	0.07	0.03	0.09	0.08	0.11
ConfAbs3	0.16	-0.15	0.1	-0.07	0.04	-0.09	0.04	0.08	0.04	-0.02	0	0.01	-0.02
ConfAbs4	0.13	-0.08	0.01	-0.09	0.06	-0.11	0.09	-0.05	-0.12	-0.02	0	0.15	0.04
ConfAbs5	0.13	-0.11	0.02	-0.02	0.02	-0.1	0.05	0.07	-0.01	0.11	-0.02	0.07	0.09
ConfAbsOverallPost	0.23**	-0.07	0.04	-0.12	0.07	-0.13	-0.02	0.13	0.07	0.03	0	0.06	0.08
ConfRelOverallPre	0.21*	-0.16	0.01	-0.19*	0.02	-0.12	-0.06	0.08	0.07	0.05	-0.04	0.18	-0.04
ConfRel1	0.22**	-0.06	0.1	-0.14	-0.02	-0.13	0.03	0.06	-0.04	0	-0.03	0.06	0.02
ConfRel2	0.16	0.05	0.12	-0.1	-0.04	-0.1	0	0.08	0.04	0.01	0.03	0.1	0.03
ConfRel3	0.13	-0.03	0.09	-0.11	0.07	-0.17	-0.04	0.04	-0.02	0.01	-0.04	0.07	0.02
ConfRel4	0.14	-0.09	0.07	-0.16	0.11	-0.18*	0.05	0.06	-0.07	0.06	0.02	0.17	0.08
ConfRel5	0.21*	-0.13	0.11	-0.03	0.07	-0.13	0.09	0.09	0.06	0.05	-0.02	0.08	0.12
ConfRelOverallPost	0.34**	-0.14	0.12	-0.13	0.08	-0.12	0.03	0.07	0.06	-0.01	-0.03	0.12	0.07
ACT_COMP_SCR	0.08	0.02	-0.15	0.02	-0.01	-0.01	-0.17	0.1	-0.02	0.11	-0.03	0.24*	0.13
ACT_ENGL_SCR	-0.02	-0.01	-0.17	0.07	-0.15	0.04	-0.11	-0.02	-0.04	0.08	-0.06	0.09	-0.05
ACT_ENGWR_SCR	-0.05	0.03	-0.31**	0.08	-0.02	0.06	-0.15	0.03	-0.03	0.06	-0.13	0.09	-0.03
ACT_MATH_SCR	0.27**	-0.01	0.02	-0.09	-0.01	-0.13	-0.09	0.17	0.07	0.05	-0.04	0.19	0.27**
ACT_READ_SCR	-0.05	0.03	-0.18	0.06	-0.08	-0.05	-0.19	0.08	-0.02	0.11	0	0.18	0
ACT_SCIRE_SCR	-0.03	0.08	-0.2	0.08	-0.03	-0.1	-0.03	0.2	0.03	0.11	-0.03	0.17	0.11
CUM_GPA	0	0	-0.04	-0.04	0.22*	-0.06	0.08	-0.15	-0.2*	-0.12	0.05	-0.04	-0.05
TOT_ACAD_HOURS	-0.1	-0.08	0.01	0.01	0.04	-0.2*	-0.12	0.13	0.3**	0.18*	0.18*	-0.41**	0
HS_RANK_PCT	0.11	-0.03	-0.02	0.07	-0.01	-0.08	-0.03	-0.15	-0.18	-0.08	-0.03	0.23	0.18

<u>Notes</u>. For a description of each variable, see <u>Table 5</u>. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	pvIPC7	nvIPC7	pemIPC7	nemIPC7	conIPC7	agreeIPC7	cnvIPC7	ReaRIASEC	InvRIASEC	ArtRIASEC	SocRIASEC	EntRIASEC	ConRIASEC
Gender	0.17	0.16	-0.05	-0.24**	-0.01	-0.22*	0.13	0.21*	-0.13	-0.15	-0.34**	0.17*	0.15
extroGoldberg	0.39**	-0.2*	0.82**	-0.18*	-0.19*	-0.24**	-0.11	-0.07	0.16	0.12	0.3**	0.21*	-0.15
neurotGoldberg	-0.09	0.05	-0.02	0.69**	0.08	-0.08	-0.1	-0.14	-0.04	0.14	0.04	-0.12	-0.11
intelGoldberg	0.22*	-0.13	0.24**	0.05	0	-0.2*	-0.39**	0.13	0.13	0.54**	0.39**	0.1	-0.06
agreeGoldberg	0.23**	-0.4**	0.4**	-0.11	0.16	0.42**	0.02	-0.15	0	0.14	0.33**	0.14	-0.08
conGoldberg	0.28**	-0.3**	0.15	-0.09	0.7**	-0.06	0.16	0.01	0.05	0	0.14	0.01	0.2*
pvIPC7	1**	-0.28**	0.44**	-0.13	0.16	-0.17	0.06	0.03	0.05	0.05	0.09	0.3**	0.07
nvIPC7	-0.28**	1**	-0.35**	0.14	-0.14	-0.19*	-0.13	0.15	-0.05	-0.01	-0.14	0.05	0.27**
pemIPC7	0.44**	-0.35**	1**	-0.09	-0.12	-0.08	-0.07	-0.1	0.16	0.15	0.36**	0.17*	-0.2*
nemIPC7	-0.13	0.14	-0.09	1**	0.09	-0.04	-0.03	-0.14	-0.07	0.12	0.05	-0.11	0.04
conIPC7	0.16	-0.14	-0.12	0.09	1**	0.02	0.31**	-0.03	-0.09	-0.12	-0.12	-0.12	0.22*
agreeIPC7	-0.17	-0.19*	-0.08	-0.04	0.02	1**	0.14	-0.22*	-0.08	-0.09	-0.07	-0.07	-0.1
cnvIPC7	0.06	-0.13	-0.07	-0.03	0.31**	0.14	1**	0.02	0.03	-0.39**	-0.13	-0.06	0.12
ReaRIASEC	0.03	0.15	-0.1	-0.14	-0.03	-0.22*	0.02	1**	0.41**	0.3**	0.14	0.13	0.45**
InvRIASEC	0.05	-0.05	0.16	-0.07	-0.09	-0.08	0.03	0.41**	1**	0.15	0.27**	-0.19*	0.03
ArtRIASEC	0.05	-0.01	0.15	0.12	-0.12	-0.09	-0.39**	0.3**	0.15	1**	0.46**	0.24**	-0.09
SocRIASEC	0.09	-0.14	0.36**	0.05	-0.12	-0.07	-0.13	0.14	0.27**	0.46**	1**	0.16	-0.02
EntRIASEC	0.3**	0.05	0.17*	-0.11	-0.12	-0.07	-0.06	0.13	-0.19*	0.24**	0.16	1**	0.23**
ConRIASEC	0.07	0.27**	-0.2*	0.04	0.22*	-0.1	0.12	0.45**	0.03	-0.09	-0.02	0.23**	1**

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	pvIPC7	nvIPC7	pemIPC7	nemIPC7	conIPC7	agreeIPC7	cnvIPC7	ReaRIASEC	InvRIASEC	ArtRIASEC	SocRIASEC	EntRIASEC	ConRIASEC
numpredout	-0.05	-0.08	0.09	-0.04	-0.02	0.06	-0.08	0.02	0.05	0.17*	0.1	0	0
numratees	-0.07	-0.12	0.04	-0.04	-0.01	0.07	0.13	-0.12	0.02	-0.03	-0.06	-0.09	-0.07
lengthtimepred	-0.2*	-0.04	-0.09	-0.04	0.1	0.03	0.07	0.06	0.13	0.04	0.02	-0.2*	0.06
numpredoutformal	-0.03	0	0.04	-0.07	-0.03	0.06	-0.1	0.1	0.01	0.09	0.03	0.02	0.07
numpredouttraining	0.08	-0.16	0.16	-0.08	0.04	0.13	-0.11	-0.11	-0.04	0.14	-0.02	0.01	-0.13
numpredouttrainingformal	-0.02	-0.12	0.12	-0.05	-0.02	0.01	0.01	-0.06	-0.05	0.13	-0.02	0.06	-0.21*
lengthtraining	-0.13	-0.06	-0.16	-0.05	-0.17*	0.04	-0.03	-0.15	-0.17	0.11	-0.05	0.09	-0.14
amtstats	-0.01	-0.1	0.1	0.03	0.02	0.09	-0.13	-0.03	0.01	0.18*	0.17*	-0.05	0.07
amtdecisionmkg	0	0.06	-0.04	-0.01	0.03	-0.07	-0.17	-0.01	-0.01	0.1	-0.11	0.02	0.12
TrPerExprAware	0.09	-0.13	-0.03	-0.05	0.07	-0.06	-0.02	0.1	0.04	-0.07	0.03	-0.02	0.18*
TrPerSaidUsed	-0.02	-0.13	-0.08	-0.06	-0.01	-0.03	-0.06	0.11	0.01	-0.04	0.06	-0.04	0.16
TrPerUsedCorrectly	-0.01	-0.04	0.03	0.05	-0.03	0	0.03	0.01	-0.06	-0.16	0	0.02	0.01
TrPerSaidUsed3rdVar	0.1	0.1	-0.1	0.13	-0.01	-0.31**	0.05	0.1	-0.04	0.09	0.07	0.17	0.13

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	numpredout	numratees	lengthtimepred	numpredoutformal	numpredouttraining	numpredouttrainingformal	lengthtraining	amtstats	amtdecisionmkg	TrPerExprAware	TrPerSaidUsed	TrPerUsedCorrectly	TrPerSaidUsed3rdVar
Fisherra	-0.08	-0.04	-0.1	-0.18*	0.01	-0.09	-0.1	-0.04	0.14	0.25**	0.29**	0.3**	-0.03
Fisherra1	-0.1	-0.12	-0.09	-0.12	-0.07	-0.08	-0.15	-0.21*	-0.1	0	-0.04	-0.11	-0.1
Fisherra2	0.01	0.03	-0.04	-0.05	0.16	-0.09	0.11	0.11	0.06	-0.02	-0.03	0.04	0.14
Fisherra3	-0.14	-0.03	0.03	-0.05	-0.04	-0.01	-0.04	0.06	0.1	0.14	0.2*	0.28**	-0.17*
Fisherra4	0.1	0.04	-0.13	-0.09	-0.02	-0.06	-0.02	0.05	0.1	0.24**	0.21*	0.33**	0.05
Fisherra5	-0.01	0	0.01	-0.1	0.03	-0.02	-0.06	-0.08	0.1	0.23**	0.28**	0.13	0.03
SS	-0.04	-0.03	-0.02	-0.08	-0.02	-0.04	-0.08	0.04	0.1	0.24**	0.32**	0.26**	0.05
SS1	-0.04	-0.02	0	-0.02	-0.11	0	-0.06	-0.12	0	0.03	0.03	-0.09	0
SS2	0.08	0.06	0.05	0.01	0.07	-0.07	0.07	0.01	0.04	0.07	0.06	0.06	0.09
SS3	-0.01	-0.01	-0.04	-0.11	-0.14	0	0.01	0.06	0.09	0.26**	0.32**	0.26**	-0.08
SS4	0.07	0.02	-0.09	-0.03	0.02	-0.01	0.03	0.1	0.09	0.26**	0.21*	0.24*	0.1
SS5	-0.1	-0.04	0.01	-0.01	0.04	0.02	-0.14	0.05	0.05	0.03	0.16	0.18	0.01
FisherRslin	-0.08	-0.02	0.27**	0	0.04	-0.09	0.03	-0.1	-0.1	-0.29**	-0.21*	-0.25**	-0.06
FisherRslin1	-0.07	-0.1	-0.05	-0.16	0.03	-0.15	-0.03	0.02	0.01	0.09	0.16	0.09	-0.05
FisherRslin2	-0.17*	-0.22**	-0.01	-0.11	0.05	-0.02	-0.04	0.1	0.14	-0.1	-0.1	-0.31**	0.05
FisherRslin3	-0.07	-0.04	0.25**	-0.05	0.03	-0.1	0.03	-0.06	-0.09	-0.17	-0.17*	-0.26**	-0.02
FisherRslin4	-0.05	-0.03	0.28**	0.04	-0.01	-0.1	0	0.02	-0.07	-0.28**	-0.2*	-0.23*	-0.17*
FisherRslin5	0.06	0.1	0.32**	0.08	0.05	0	0.04	-0.09	-0.15	-0.3**	-0.27**	-0.21*	-0.09

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	numpredout	numratees	lengthtimepred	numpredoutformal	numpredouttraining	numpredouttrainingformal	lengthtraining	amtstats	amtdecisionmkg	TrPerExprAware	TrPerSaidUsed	TrPerUsedCorrectly	TrPerSaidUsed3rdVar
FisherG	0.04	0.03	0.04	-0.17*	-0.05	-0.1	-0.03	0.09	-0.05	-0.12	-0.06	-0.1	0
FisherG1	-0.11	-0.07	-0.04	-0.08	-0.04	-0.05	-0.08	-0.19*	-0.09	-0.02	-0.09	-0.14	-0.09
FisherG2	-0.08	0	-0.01	-0.09	0.01	-0.16	-0.02	0.07	0.05	0.01	-0.04	-0.02	0.22*
FisherG3	-0.2*	0.01	0.01	0.02	0	0.15	-0.05	-0.01	0.1	-0.07	-0.02	0	-0.09
FisherG4	0.02	0.08	-0.05	0.15	0.13	0.06	0.1	0.07	0.03	-0.02	0.01	0	-0.01
FisherG5	-0.03	0.04	0.1	-0.01	0.06	0.01	0.02	-0.12	-0.06	0.04	0.07	-0.04	-0.02
FisherCstat	-0.04	-0.05	-0.11	-0.09	0.02	-0.07	-0.14	-0.01	0.16	0.29**	0.33**	0.36**	-0.05
FisherCstat1	-0.17*	-0.12	-0.08	0.03	0	0.02	-0.15	-0.13	0.05	-0.05	-0.06	-0.12	-0.13
FisherCstat2	0.06	0.11	0	0.01	0.11	-0.03	0.03	0.05	0.05	0.05	0.12	0.31**	0.06
FisherCstat3	0.01	0	0.07	-0.01	0.01	-0.11	-0.04	-0.03	0.16	0.16	0.22**	0.39**	-0.07
FisherCstat4	0.03	-0.14	-0.21*	-0.12	-0.05	-0.1	-0.11	0.01	0.09	0.32**	0.31**	0.34**	0.03
FisherCstat5	-0.04	-0.03	-0.09	-0.11	-0.08	-0.17*	-0.05	-0.06	0.16	0.33**	0.37**	0.18	0.01
Fisherrz	-0.04	-0.05	-0.1	-0.09	0.02	-0.07	-0.14	-0.01	0.16	0.29**	0.33**	0.36**	-0.05
Fisherrz1	-0.18*	-0.11	-0.07	0.04	0.02	0.04	-0.16	-0.13	0.07	-0.04	-0.04	-0.11	-0.14
Fisherrz2	0.05	0.12	0	0	0.06	-0.01	0	0.03	0.02	0.07	0.14	0.32**	0.04
Fisherrz3	0.04	0	0.03	-0.02	0.01	-0.13	-0.03	-0.02	0.18*	0.14	0.21*	0.37**	-0.06
Fisherrz4	0.03	-0.16	-0.2*	-0.11	-0.02	-0.09	-0.1	0.01	0.07	0.31**	0.29**	0.3**	0.03
Fisherrz5	-0.07	-0.03	-0.1	-0.11	-0.08	-0.17*	-0.04	-0.04	0.13	0.31**	0.34**	0.18	0.01

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	numpredout	numratees	lengthtimepred	numpredoutformal	numpredouttraining	numpredouttrainingformal	lengthtraining	amtstats	amtdecisionmkg	TrPerExprAware	TrPerSaidUsed	TrPerUsedCorrectly	TrPerSaidUsed3rdVar
Cwtx1x2_s	-0.02	0.05	-0.03	-0.07	-0.04	-0.14	-0.04	-0.05	0.11	0.23**	0.31**	0.31**	0.01
Cwtx1x2_1_s	-0.27**	-0.13	-0.09	-0.28**	-0.12	-0.23**	0.03	-0.14	0.09	0.01	0	0.05	0
Cwtx1x2_2_s	0.07	-0.01	-0.02	-0.12	0.01	-0.1	0.04	0.02	0.05	0.19*	0.22*	0.24*	0.02
Cwtx1x2_3_s	0.11	-0.01	-0.04	-0.1	0.02	-0.09	-0.04	0.05	0.07	0.29**	0.31**	0.24**	-0.07
Cwtx1x2_4_s	0.08	-0.05	-0.11	-0.09	0.03	-0.07	-0.08	0.1	0.15	0.19*	0.2*	0.2*	-0.12
Cwtx1x2_5_s	0	-0.06	-0.1	-0.15	-0.04	-0.16	-0.05	-0.01	0.13	0.29**	0.31**	0.19*	0.08
ConfAbsOverallPre	0.2*	0.04	0.06	0.05	0.07	-0.04	0	0.04	0	-0.01	0	-0.07	0.08
ConfAbs1	0.18*	0.07	0.11	0.06	0.06	-0.01	-0.07	0.02	0.05	-0.06	-0.02	0.04	0.01
ConfAbs2	0.15	0.05	0.11	0.04	-0.02	0.01	-0.19*	-0.01	0.05	0.07	0.12	0.2*	0.01
ConfAbs3	0.1	0.15	0.13	0.03	0.02	-0.02	0.13	-0.02	0.01	0.01	0.04	0.15	0.1
ConfAbs4	0.11	0.07	0.08	-0.04	-0.13	-0.09	0.11	-0.07	0.13	0.05	0.07	0.11	0.06
ConfAbs5	0.13	0.08	0.16	0.02	0.04	-0.02	0.03	0.01	0.14	0.07	0.1	0.09	0.07
ConfAbsOverallPost	0.16	0.1	0.15	0.04	0.05	-0.03	0.06	-0.07	0.05	0.03	0.06	0.17	0.08
ConfRelOverallPre	0.22*	0.01	0.08	0.02	0.09	-0.01	0.04	0.04	-0.01	-0.01	0.05	-0.02	-0.03
ConfRel1	0.24**	-0.02	0.05	0	0.03	-0.06	0.05	0.03	0.08	-0.11	-0.08	0.02	-0.03
ConfRel2	0.25**	0.04	0.04	0.03	0.01	0.01	-0.12	0.02	0.11	-0.01	0.07	0.19*	-0.08
ConfRel3	0.11	0.08	0.07	-0.01	0.09	-0.03	0.07	0.02	0.07	-0.06	-0.01	0.1	0.08
ConfRel4	0.21*	0.11	0.11	-0.05	-0.03	-0.1	0.15	-0.06	0.13	-0.02	0.02	0.05	0.09
ConfRel5	0.11	0.04	0.1	-0.01	0.06	-0.05	-0.02	0	0.09	0.05	0.07	0.11	0.03
ConfRelOverallPost	0.09	0.01	0.04	-0.07	0.14	-0.01	-0.01	-0.04	-0.01	0.03	0.04	0.15	0.09

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	numpredout	numratees	lengthtimepred	numpredoutformal	numpredouttraining	numpredouttrainingformal	lengthtraining	amtstats	amtdecisionmkg	TrPerExprAware	TrPerSaidUsed	TrPerUsedCorrectly	TrPerSaidUsed3rdVar
ACT_COMP_SCR	-0.02	-0.2*	-0.05	-0.07	0.02	-0.14	-0.07	-0.1	-0.09	0.06	0.11	0.04	0
ACT_ENGL_SCR	0.04	-0.11	0.04	-0.12	-0.04	-0.12	0.02	-0.13	-0.12	0.04	0.07	0.07	0.03
ACT_ENGWR_SCR	0.02	-0.03	0.17	-0.04	0.02	-0.04	0	-0.13	-0.1	0.09	0.1	0.15	0.04
ACT_MATH_SCR	-0.08	-0.14	-0.15	-0.08	-0.05	-0.05	-0.07	0	0.03	0.05	0.13	0.05	0.01
ACT_READ_SCR	-0.02	-0.18	0.02	-0.05	-0.04	-0.17	-0.04	-0.06	-0.07	-0.01	0.07	0.02	0.03
ACT_SCIRE_SCR	-0.17	-0.29**	-0.04	-0.15	-0.06	-0.09	-0.16	-0.13	-0.13	0.01	0.03	0	-0.06
CUM_GPA	-0.14	-0.23**	-0.1	-0.08	-0.17*	-0.16	-0.05	-0.11	0	0.03	0.01	-0.01	0
TOT_ACAD_HOURS	0.06	0.02	0.21*	0.08	0.13	0.09	0.05	0.27**	-0.02	0.07	0.06	-0.12	-0.05
HS_RANK_PCT	-0.02	-0.3**	-0.37**	0.02	-0.1	-0.14	0.01	0.11	0.12	-0.03	0.06	-0.04	0.13
Gender	-0.04	0.01	-0.08	-0.03	0.01	0.1	-0.09	-0.2*	0.03	0.06	0.04	0.09	0.04
extroGoldberg	0.07	0.01	-0.07	-0.01	0.1	0.07	-0.1	0.04	0.03	-0.02	-0.06	-0.03	-0.04
neurotGoldberg	-0.05	-0.03	-0.05	0.01	-0.07	-0.03	-0.02	-0.03	0.03	-0.03	-0.05	0.15	0.01
intelGoldberg	0.13	-0.11	0.14	-0.08	0.13	0.02	0.04	0.17*	0.1	0.05	0.06	-0.1	0.06
agreeGoldberg	0.1	0.02	-0.02	0.06	0.16	0.08	-0.11	0.2*	-0.05	0.04	0.06	0.04	-0.1
conGoldberg	0.08	-0.02	0.15	0.02	0.06	-0.07	-0.07	0.11	0.05	0.15	0.07	-0.02	0.05
pvIPC7	-0.05	-0.07	-0.2*	-0.03	0.08	-0.02	-0.13	-0.01	0	0.09	-0.02	-0.01	0.1
nvIPC7	-0.08	-0.12	-0.04	0	-0.16	-0.12	-0.06	-0.1	0.06	-0.13	-0.13	-0.04	0.1
pemIPC7	0.09	0.04	-0.09	0.04	0.16	0.12	-0.16	0.1	-0.04	-0.03	-0.08	0.03	-0.1
nemIPC7	-0.04	-0.04	-0.04	-0.07	-0.08	-0.05	-0.05	0.03	-0.01	-0.05	-0.06	0.05	0.13
conIPC7	-0.02	-0.01	0.1	-0.03	0.04	-0.02	-0.17*	0.02	0.03	0.07	-0.01	-0.03	-0.01
agreeIPC7	0.06	0.07	0.03	0.06	0.13	0.01	0.04	0.09	-0.07	-0.06	-0.03	0	-0.31**
cnvIPC7	-0.08	0.13	0.07	-0.1	-0.11	0.01	-0.03	-0.13	-0.17	-0.02	-0.06	0.03	0.05

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 19 – *cont'd.*

	numpredout	numratees	lengthtimepred	numpredoutformal	numpredouttraining	numpredouttrainingformal	lengthtraining	amtstats	amtdecisionmkg	TrPerExprAware	TrPerSaidUsed	TrPerUsedCorrectly	TrPerSaidUsed3rdVar
ReaRIASEC	0.02	-0.12	0.06	0.1	-0.11	-0.06	-0.15	-0.03	-0.01	0.1	0.11	0.01	0.1
InvRIASEC	0.05	0.02	0.13	0.01	-0.04	-0.05	-0.17	0.01	-0.01	0.04	0.01	-0.06	-0.04
ArtRIASEC	0.17*	-0.03	0.04	0.09	0.14	0.13	0.11	0.18*	0.1	-0.07	-0.04	-0.16	0.09
SocRIASEC	0.1	-0.06	0.02	0.03	-0.02	-0.02	-0.05	0.17*	-0.11	0.03	0.06	0	0.07
EntRIASEC	0	-0.09	-0.2*	0.02	0.01	0.06	0.09	-0.05	0.02	-0.02	-0.04	0.02	0.17
ConRIASEC	0	-0.07	0.06	0.07	-0.13	-0.21*	-0.14	0.07	0.12	0.18*	0.16	0.01	0.13
numpredout	1**	0.36**	0.23**	0.29**	0.23**	0.11	0.26**	0.27**	0.03	0.07	0.07	0.07	0.03
numratees	0.36**	1**	0.36**	0.26**	0.22**	0.36**	0.38**	0.11	-0.08	-0.06	0.01	0.12	0.07
lengthtimepred	0.23**	0.36**	1**	0.23**	0.09	-0.02	0.11	-0.11	-0.08	0.04	0.07	0.07	0.02
numpredoutformal	0.29**	0.26**	0.23**	1**	0.44**	0.38**	-0.03	0.07	0.02	-0.05	-0.02	0.02	-0.09
numpredouttraining	0.23**	0.22**	0.09	0.44**	1**	0.62**	0.03	0.14	-0.08	-0.07	-0.09	0	-0.06
numpredouttrainingformal	0.11	0.36**	-0.02	0.38**	0.62**	1**	-0.01	0.11	-0.06	-0.08	-0.13	0	-0.07
lengthtraining	0.26**	0.38**	0.11	-0.03	0.03	-0.01	1**	0.07	-0.03	-0.13	-0.11	-0.12	0.37**
amtstats	0.27**	0.11	-0.11	0.07	0.14	0.11	0.07	1**	0.06	-0.03	-0.01	-0.01	0.01
amtdecisionmkg	0.03	-0.08	-0.08	0.02	-0.08	-0.06	-0.03	0.06	1**	0	0.04	-0.01	-0.07
TrPerExprAware	0.07	-0.06	0.04	-0.05	-0.07	-0.08	-0.13	-0.03	0	1**	0.86**	0.58**	0.06
TrPerSaidUsed	0.07	0.01	0.07	-0.02	-0.09	-0.13	-0.11	-0.01	0.04	0.86**	1**	0.59**	0.08
TrPerUsedCorrectly	0.07	0.12	0.07	0.02	0	0	-0.12	-0.01	-0.01	0.58**	0.59**	1**	0.09
TrPerSaidUsed3rdVar	0.03	0.07	0.02	-0.09	-0.06	-0.07	0.37**	0.01	-0.07	0.06	0.08	0.09	1**

Notes. For a description of each variable, see Table 5. Reported above are Pearson Product moment correlations. ** = $p < 0.01$ * = $p < 0.05$ The presence of any stars (asterisks) indicates that the confidence interval for the correlation does not contain 0.

Table 20. *Incremental Prediction of Change Over Time in Judgment Validity (ra): Individual Differences Variables (Simpler Model = Fixed Effects Only, Time x Feedback (Spring 2010)*

		Fix Eff Ind Diff	Std Err Fix Eff	t Fix Eff	pval of t	sig of t	res df simpler model	res df more complex model	RSS simpler model	RSS more complex model	F	pval of F	sig of F
1	**ACT_COMP_SCR**	0.008	0.004	1.981	0.048	*	494	493	55.724	55.284	3.925	0.048	*
2	**ACT_ENGL_SCR**	0.005	0.003	1.405	0.161		474	473	53.138	52.918	1.974	0.161	
3	**ACT_ENGWR_SCR**	0.008	0.004	1.796	0.073		419	418	47.578	47.214	3.226	0.073	
4	**ACT_MATH_SCR**	0.008	0.003	2.961	0.003	**	474	473	53.138	52.172	8.765	0.003	**
5	**ACT_READ_SCR**	0.003	0.003	0.800	0.424		474	473	53.138	53.067	0.640	0.424	
6	**ACT_SCIRE_SCR**	0.005	0.004	1.241	0.215		474	473	53.138	52.966	1.540	0.215	
7	**CUM_GPA**	0.038	0.027	1.429	0.154		659	658	73.634	73.407	2.042	0.154	
8	**TOT_ACAD_HOURS**	0.000	0.000	-0.353	0.724		659	658	73.634	73.620	0.125	0.724	
9	**HS_RANK_PCT**	0.001	0.001	1.259	0.209		394	393	43.283	43.109	1.586	0.209	
10	**Gender**	0.044	0.027	1.640	0.101		684	683	80.249	79.934	2.690	0.101	
11	**extroGoldberg**	0.001	0.001	1.096	0.273		659	658	78.683	78.540	1.201	0.273	
12	**neurotGoldberg**	0.000	0.001	-0.701	0.484		659	658	78.683	78.624	0.491	0.484	
13	**intelGoldberg**	0.001	0.001	1.365	0.173		659	658	78.683	78.461	1.862	0.173	
14	**agreeGoldberg**	0.000	0.001	0.551	0.582		659	658	78.683	78.647	0.304	0.582	
15	**conGoldberg**	0.001	0.001	1.707	0.088		659	658	78.683	78.336	2.915	0.088	

Notes. Fix Eff = fixed effect; Ind Diff = individual difference; Std Err = standard error; t Fix Eff = t-value of the fixed effect; pval of t = probability value for the t-value; sig of t = statistical significance indicator for t-value; res df = residual degreesof freedom; simpler model = regression model that best fits the data without the individual difference variable included; more complex model = simpler model with the individualdifference included; RSS = residual sum of squares; F = F-value for test of simpler model fit versus more complex model fit; pval of F = probability value of F; sig of F = statistical significance indicator for F-value; ** = p-value < 0.01; * = p-value < 0.05.

Table 20 – *cont'd.*

		Fix Eff Ind Diff	Std Err Fix Eff	t Fix Eff	pval of t	sig of t	res df simpler model	res df more complex model	RSS simpler model	RSS more complex model	F	pval of F	sig of F
16	**pvIPC7**	0.004	0.003	1.587	0.113		659	658	78.683	78.383	2.518	0.113	
17	**nvIPC7**	-0.004	0.004	-1.074	0.283		659	658	78.683	78.545	1.154	0.283	
18	**pemIPC7**	0.001	0.002	0.579	0.563		659	658	78.683	78.643	0.335	0.563	
19	**nemIPC7**	-0.003	0.004	-0.761	0.447		659	658	78.683	78.614	0.579	0.447	
20	**conIPC7**	0.004	0.003	1.525	0.128		659	658	78.683	78.406	2.324	0.128	
21	**agreeIPC7**	0.002	0.004	0.482	0.630		659	658	78.683	78.655	0.233	0.630	
22	**cnvIPC7**	-0.001	0.003	-0.503	0.615		659	658	78.683	78.653	0.253	0.615	
23	**ReaRIASEC**	0.000	0.002	-0.213	0.831		659	658	78.683	78.678	0.045	0.831	
24	**InvRIASEC**	-0.001	0.001	-0.453	0.651		659	658	78.683	78.659	0.205	0.651	
25	**ArtRIASEC**	-0.001	0.002	-0.382	0.703		659	658	78.683	78.666	0.146	0.703	
26	**SocRIASEC**	-0.002	0.002	-1.338	0.181		659	658	78.683	78.470	1.791	0.181	
27	**EntRIASEC**	0.000	0.002	0.037	0.971		659	658	78.683	78.683	0.001	0.971	
28	**ConRIASEC**	-0.001	0.002	-0.344	0.731		659	658	78.683	78.669	0.119	0.731	
29	**numpredout**	-0.005	0.011	-0.468	0.640		689	688	80.865	80.839	0.219	0.640	
30	**numratees**	0.000	0.000	0.083	0.934		689	688	80.865	80.864	0.007	0.934	
31	**lengthtimepred**	0.000	0.000	-1.164	0.245		689	688	80.865	80.706	1.356	0.245	
32	**numpredoutformal**	-0.048	0.026	-1.800	0.072		689	688	80.865	80.486	3.240	0.072	
33	**numpredouttraining**	0.017	0.033	0.502	0.616		689	688	80.865	80.835	0.252	0.616	
34	**numpredouttrainingformal**	-0.050	0.057	-0.881	0.379		689	688	80.865	80.774	0.776	0.379	
35	**lengthtraining**	0.000	0.000	-0.782	0.434		684	683	80.159	80.088	0.612	0.434	
36	**amtstats**	0.001	0.024	0.058	0.953		694	693	82.071	82.071	0.003	0.953	
37	**amtdecisionmkg**	0.084	0.056	1.493	0.136		694	693	82.071	81.808	2.230	0.136	

Notes. Fix Eff = fixed effect; Ind Diff = individual difference; Std Err = standard error; t Fix Eff = t-value of the fixed effect; pval of t = probability value for the t-value; sig of t = statistical significance indicator for t-value; res df = residual degreesof freedom; simpler model = regression model that best fits the data without the individual difference variable included; more complex model = simpler model with the individualdifference included; RSS = residual sum of squares; F = F-value for test of simpler model fit versus more complex model fit; pval of F = probability value of F; sig of F = statistical significance indicator for F-value; ** = p-value < 0.01; * = p-value < 0.05.

Table 21. *Incremental Prediction of Change Over Time in Skill Score: Individual Differences Variables (Simpler Model = Fixed Effects Only, No Time x Feedback Interaction) (Spring 2010)*

		Fix Eff Ind Diff	Std Err Fix Eff	t Fix Eff	pval of t	sig of t	res df simpler model	res df more complex model	RSS simpler model	RSS more complex model	F	pval of F	sig of F
1	**ACT_COMP_SCR**	0.044	0.012	3.754	0.000	**	498	497	468.712	455.789	14.091	0.000	**
2	**ACT_ENGL_SCR**	0.031	0.009	3.374	0.001	**	478	477	441.087	430.804	11.386	0.001	**
3	**ACT_ENGWR_SCR**	0.038	0.012	3.019	0.003	**	423	422	418.634	409.781	9.116	0.003	**
4	**ACT_MATH_SCR**	0.035	0.008	4.313	0.000	**	478	477	441.087	424.528	18.606	0.000	**
5	**ACT_READ_SCR**	0.028	0.009	3.070	0.002	**	478	477	441.087	432.541	9.424	0.002	**
6	**ACT_SCIRE_SCR**	0.033	0.013	2.661	0.008	**	478	477	441.087	434.634	7.082	0.008	**
7	**CUM_GPA**	0.097	0.116	0.838	0.402		663	662	1403.027	1401.539	0.703	0.402	
8	**TOT_ACAD_HOURS**	0.001	0.002	0.533	0.594		663	662	1403.027	1402.425	0.284	0.594	
9	**HS_RANK_PCT**	0.005	0.002	2.393	0.017	*	398	397	376.008	370.662	5.726	0.017	*
10	**Gender**	0.020	0.110	0.183	0.855		688	687	1413.322	1413.252	0.034	0.855	
11	**extroGoldberg**	0.002	0.003	0.649	0.516		663	662	1421.874	1420.969	0.422	0.516	
12	**neurotGoldberg**	-0.004	0.003	-1.300	0.194		663	662	1421.874	1418.252	1.691	0.194	
13	**intelGoldberg**	0.006	0.003	1.729	0.084		663	662	1421.874	1415.479	2.991	0.084	
14	**agreeGoldberg**	0.003	0.003	0.819	0.413		663	662	1421.874	1420.435	0.671	0.413	
15	**conGoldberg**	0.002	0.003	0.701	0.484		663	662	1421.874	1420.821	0.491	0.484	

Notes. Fix Eff = fixed effect; Ind Diff = individual difference; Std Err = standard error; t Fix Eff = t-value of the fixed effect; pval of t = probability value for the t-value; sig of t = statistical significance indicator for t-value; res df = residual degreesof freedom; simpler model = regression model that best fits the data without the individual difference variable included; more complex model = simpler model with the individualdifference included; RSS = residual sum of squares; F = F-value for test of simpler model fit versus more complex model fit; pval of F = probability value of F; sig of F = statistical significance indicator for F-value; ** = p-value < 0.01; * = p-value < 0.05.

Table 21 – *cont'd.*

		Fix Eff Ind Diff	Std Err Fix Eff	t Fix Eff	pval of t	sig of t	res df simpler model	res df more complex model	RSS simpler model	RSS more complex model	F	pval of F	sig of F
16	**pvIPC7**	0.005	0.011	0.458	0.647		663	662	1421.874	1421.423	0.210	0.647	
17	**nvIPC7**	-0.027	0.015	-1.791	0.074		663	662	1421.874	1415.014	3.209	0.074	
18	**pemIPC7**	0.005	0.009	0.579	0.563		663	662	1421.874	1421.154	0.335	0.563	
19	**nemIPC7**	0.002	0.015	0.156	0.876		663	662	1421.874	1421.822	0.024	0.876	
20	**conIPC7**	0.001	0.011	0.102	0.919		663	662	1421.874	1421.852	0.010	0.919	
21	**agreeIPC7**	0.003	0.015	0.210	0.834		663	662	1421.874	1421.780	0.044	0.834	
22	**cnvIPC7**	-0.012	0.012	-1.019	0.308		663	662	1421.874	1419.646	1.039	0.308	
23	**ReaRIASEC**	0.002	0.007	0.352	0.725		663	662	1421.874	1421.609	0.124	0.725	
24	**InvRIASEC**	-0.001	0.006	-0.109	0.913		663	662	1421.874	1421.849	0.012	0.913	
25	**ArtRIASEC**	0.008	0.007	1.201	0.230		663	662	1421.874	1418.783	1.443	0.230	
26	**SocRIASEC**	0.001	0.008	0.128	0.899		663	662	1421.874	1421.839	0.016	0.899	
27	**EntRIASEC**	-0.005	0.007	-0.680	0.497		663	662	1421.874	1420.881	0.463	0.497	
28	**ConRIASEC**	-0.012	0.007	-1.815	0.070		663	662	1421.874	1414.834	3.294	0.070	
29	**numpredout**	-0.025	0.047	-0.525	0.600		693	692	1444.839	1444.265	0.275	0.600	
30	**numratees**	0.000	0.002	-0.190	0.849		693	692	1444.839	1444.764	0.036	0.849	
31	**lengthtimepred**	0.000	0.001	-0.220	0.826		693	692	1444.839	1444.738	0.048	0.826	
32	**numpredoutformal**	-0.075	0.108	-0.688	0.492		693	692	1444.839	1443.851	0.473	0.492	
33	**numpredouttraining**	-0.031	0.139	-0.224	0.823		693	692	1444.839	1444.734	0.050	0.823	
34	**numpredouttrainingformal**	-0.046	0.235	-0.194	0.846		693	692	1444.839	1444.760	0.038	0.846	
35	**lengthtraining**	-0.001	0.001	-1.033	0.302		688	687	1442.040	1439.803	1.067	0.302	
36	**amtstats**	0.056	0.097	0.573	0.567		698	697	1456.093	1455.406	0.329	0.567	
37	**amtdecisionmkg**	0.289	0.235	1.229	0.219		698	697	1456.093	1452.942	1.511	0.219	

Notes. Fix Eff = fixed effect; Ind Diff = individual difference; Std Err = standard error; t Fix Eff = t-value of the fixed effect; pval of t = probability value for the t-value; sig of t = statistical significance indicator for t-value; res df = residual degrees of freedom; simpler model = regression model that best fits the data without the individual difference variable included; more complex model = simpler model with the individualdifference included; RSS = residual sum of squares; F = F-value for test of simpler model fit versus more complex model fit; pval of F = probability value of F; sig of F = statistical significance indicator for F-value; ** = p-value < 0.01; * = p-value < 0.05.

Table 22. *Insight: Interrater Agreement (Spring 2010)*

Block	Dimension Rated	Disagreement (Absolute Frequencies)					Agreement (Relative Frequencies)			
		Overall	C1	C2	C3		Overall	C1	C2	C3
1	Expressed awareness of disordinal interaction?	38	12	12	14		69%	70%	71%	66%
	Said s/he used disordinal interaction?	31	12	8	11		75%	70%	81%	73%
	Used disordinal interaction properly?	22	6	7	9		82%	85%	83%	78%
2	Expressed awareness of disordinal interaction?	25	11	8	6		80%	73%	81%	85%
	Said s/he used disordinal interaction?	21	11	5	5		83%	73%	88%	88%
	Used disordinal interaction properly?	20	10	5	5		84%	75%	88%	88%
3	Expressed awareness of disordinal interaction?	25	13	8	4		80%	68%	81%	90%
	Said s/he used disordinal interaction?	20	12	5	3		84%	70%	88%	93%
	Used disordinal interaction properly?	25	10	6	9		80%	75%	86%	78%
4	Expressed awareness of disordinal interaction?	20	8	8	4		84%	80%	81%	90%
	Said s/he used disordinal interaction?	17	9	5	3		86%	78%	88%	93%
	Used disordinal interaction properly?	21	7	6	8		83%	83%	86%	80%
5	Expressed awareness of disordinal interaction?	13	5	5	3		89%	88%	88%	93%
	Said s/he used disordinal interaction?	10	5	2	3		92%	88%	95%	93%
	Used disordinal interaction properly?	18	5	5	8		85%	88%	88%	80%
Overall (after all predictions were made)	With which block did s/he think his/her strategy changed (no change = 0)?	14	6	3	5		89%	85%	93%	88%
	Said s/he used disordinal interaction?	9	5	0	4		93%	88%	100%	90%
	Used disordinal interaction properly?	14	7	2	5		89%	83%	95%	88%

Notes. C1, C2, and C3 represent the 3 different feedback conditions. All coding was binary (“yes” or “no”), so each disagreement = + 1.

Table 23. *Incremental Prediction of Change Over Time in Judgment Validity (r_a): Insight (Simpler Model = Fixed Effects Only, Time x Feedback Interaction) (Spring 2010)*

		Fix Eff Ind Diff	Std Err Fix Eff	t Fix Eff	pval of t	sig of t	res df simpler model	res df more complex model	RSS simpler model	RSS more complex model	F	pval of F	sig of F
1	**TrPerExprAware**	0.043	0.015	2.834	0.005	**	669	668	79.779	78.832	8.030	0.005	**
2	**TrPerSaidUsed**	0.042	0.016	2.696	0.007	**	669	668	79.779	78.921	7.266	0.007	**
3	**TrPerUsedCorrectly**	0.045	0.014	3.075	0.002	**	559	558	67.118	65.999	9.458	0.002	**

Table 24. *Incremental Prediction of Change Over Time in Skill Score: Insight (Simpler Model = Fixed Effects Only, No Time x Feedback Interaction) (Spring 2010)*

		Fix Eff Ind Diff	Std Err Fix Eff	t Fix Eff	pval of t	sig of t	res df simpler model	res df more complex model	RSS simpler model	RSS more complex model	F	pval of F	sig of F
1	**TrPerExprAware**	0.167	0.064	2.623	0.009	**	673	672	1432.502	1417.987	6.879	0.009	**
2	**TrPerSaidUsed**	0.239	0.064	3.755	0.000	**	673	672	1432.502	1403.065	14.099	0.000	**
3	**TrPerUsedCorrectly**	0.173	0.060	2.881	0.004	**	563	562	1169.593	1152.565	8.303	0.004	**

Notes. Fix Eff = fixed effect; Ind Diff = individual difference; Std Err = standard error; t Fix Eff = t-value of the fixed effect; pval of t = probability value for the t-value; sig of t = statistical significance indicator for t-value; res df = residual degreesof freedom; simpler model = regression model that best fits the data without the individual difference variable included; more complex model = simpler model with the individual difference included; RSS = residual sum of squares; F = F-value for test of simpler model fit versus more complex model fit; pval of F = probability value of F; sig of F = statistical significance indicator for F-value; ** = p-value < 0.01; * = p-value < 0.05.

Table 25. *Insight: Dimensions Coded Based on Narrative Self-Reports (Spring 2010)*

When Insight Is Achieved	Aware of Disordinal Interaction					
	All Reversals Ignored		Any Reversal → Insight Indeterminate		Any Reversal → No Insight	
	N	%	*N*	%	*N*	%
Never	30	21%	30	26%	56	40%
After block 1	46	33%	35	31%	35	25%
After block 2	38	27%	29	25%	29	21%
After block 3	16	11%	11	10%	11	8%
After block 4	9	6%	8	7%	8	6%
After block 5	1	1%	1	1%	1	1%
Total	140	100%	114	100%	140	100%

When Insight Is Achieved	Used Disordinal Interaction					
	All Reversals Ignored		Any Reversal → Insight Indeterminate		Any Reversal → No Insight	
	N	%	*N*	%	*N*	%
Never	31	22%	31	31%	70	50%
After block 1	42	30%	27	27%	27	19%
After block 2	41	29%	26	26%	26	19%
After block 3	16	12%	9	9%	9	6%
After block 4	7	5%	5	5%	5	4%
After block 5	2	1%	2	2%	2	1%
Total	139	100%	100	100%	139	100%

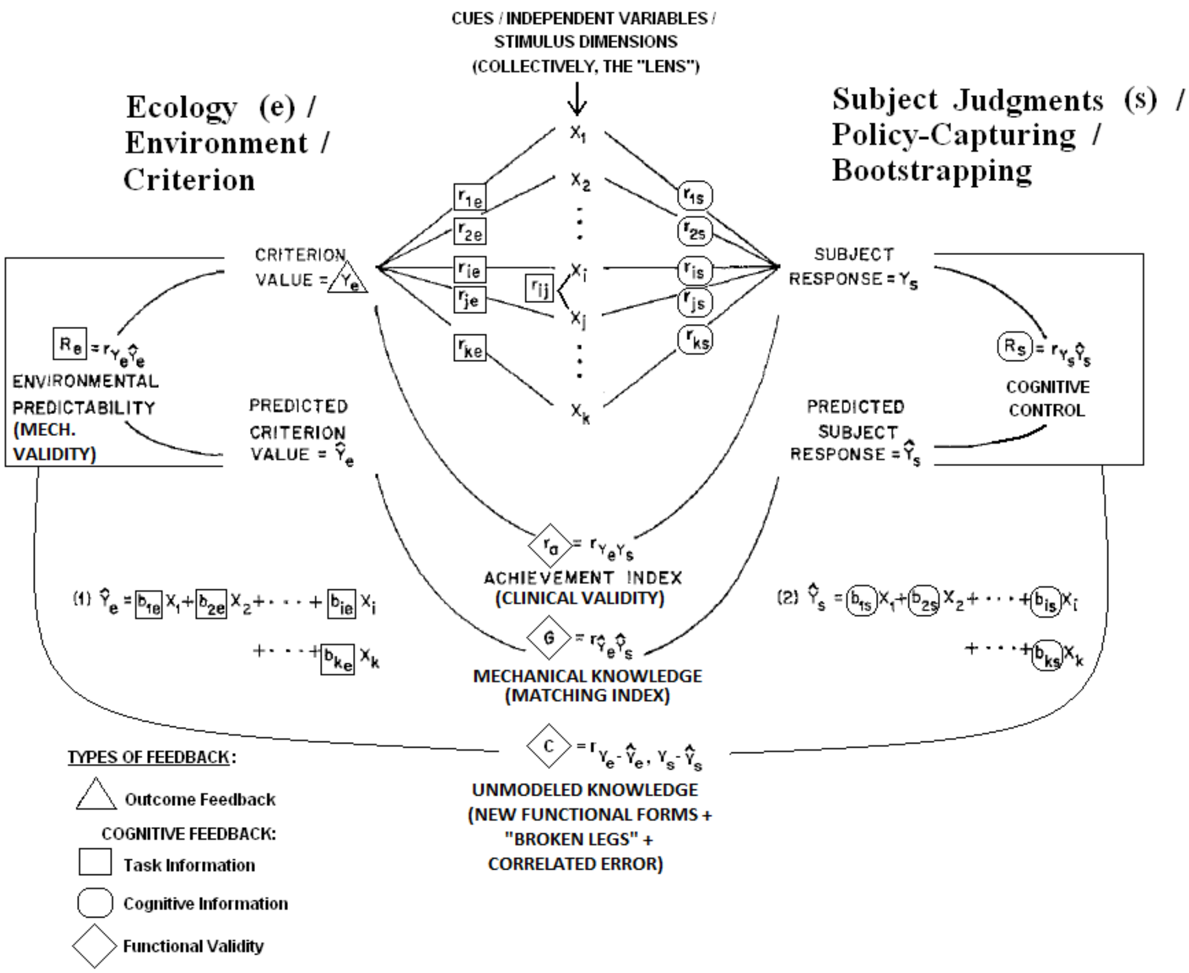
When Insight Is Achieved	Properly Used Disordinal Interaction					
	All Reversals Ignored		Any Reversal → Insight Indeterminate		Any Reversal → No Insight	
	N	%	*N*	%	*N*	%
Never	21	20%	21	23%	36	34%
After block 1	48	45%	39	42%	39	36%
After block 2	16	15%	14	15%	14	13%
After block 3	15	14%	12	13%	12	11%
After block 4	4	4%	3	3%	3	3%
After block 5	3	3%	3	3%	3	3%
Total	107	100%	92	100%	107	100%

Notes. "All Reversals Ignored" = insight is achieved even if a reversal later occurs; "Any Reversal → Insight Indeterminate" = subject is removed from the analysis if there is any reversal; "Any Reversal → No Insight" = subject is considered to have not gained insight if there is any reversal; *N* = number of subjects; % = percentage of subjects (based on *N*s)

Appendix S

FIGURES

Figure 1. Lens Model Diagram



(reproduced with modification from Slovic & Lichtenstein, 1971 and Balzer et al., 1989)

Note that "MECH. VALIDITY" and "CLINICAL VALIDITY" (above) refer to criterion-related validity (accuracy of predicting the criterion) based on a correlation metric.

<u>The Lens Model Equation</u>:

$$r_a = r_{Ye,Ys} = GR_eR_s + C\sqrt{(1-R_e^2)}\sqrt{(1-R_s^2)}$$

See <u>Table 1</u> for the variables' definitions and descriptions.

Figure 2. Relationship of Cognitive Control (R_s) to Number of Persons Assessed in Naturalistic Environments

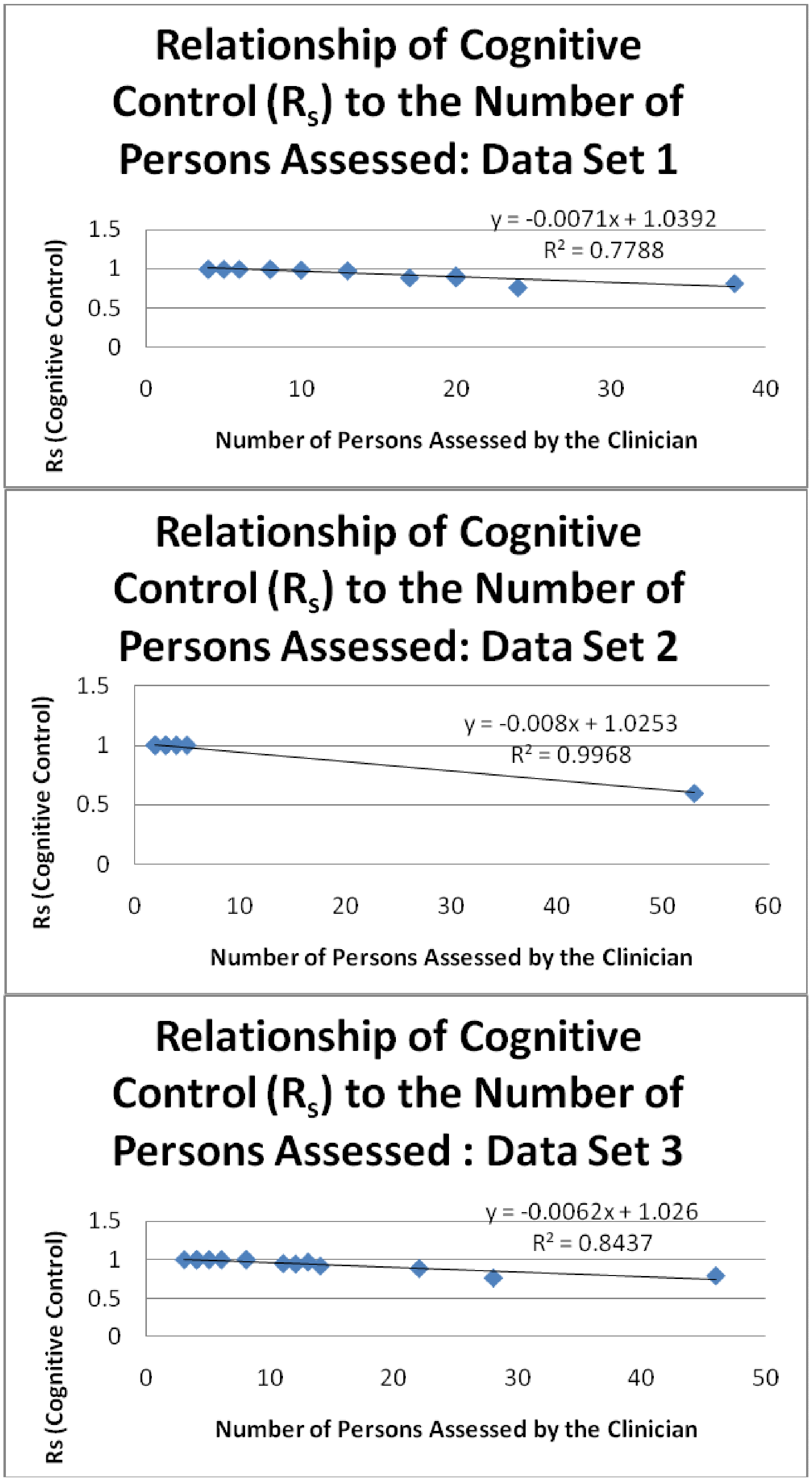
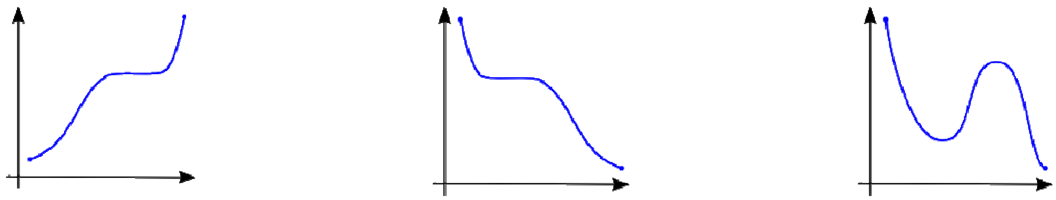


Figure 3. Examples of Monotonic and Non-Monotonic Relationships



Monotonic Relationship Between X and Y
The line is continually non-decreasing (the slope or slant of the line at all points is either positive or 0).

Monotonic Relationship Between X and Y
The line is continually non-increasing (the slope or slant of the line at all points is either negative or 0).

Not a Monotonic Relationship Between X and Y
The line is not continually non-decreasing or continually non-increasing, but instead it changes direction (goes from decrease to increase or from increase to decrease) at least once.

Illustrations based on http://en.wikipedia.org/wiki/Monotonic_function (accessed January 11, 2009)

Figure 4. Curved Logarithmic (and Monotone) Function Approximated by a Straight Line

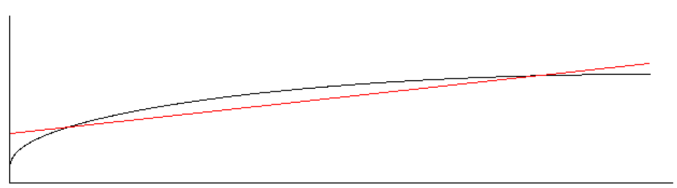
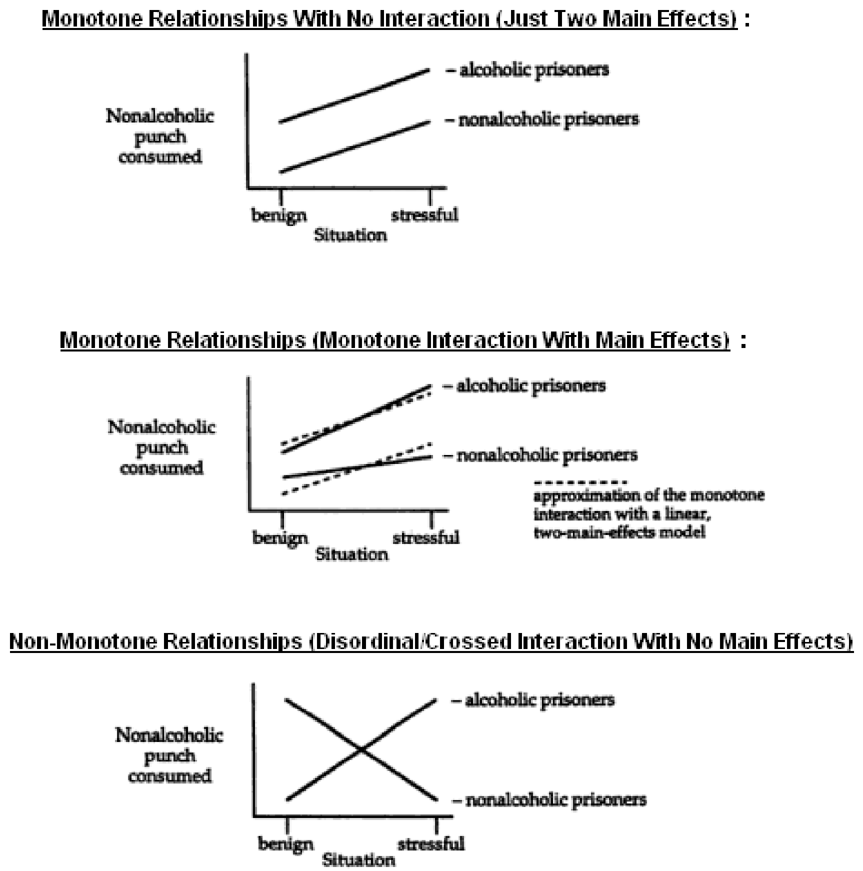
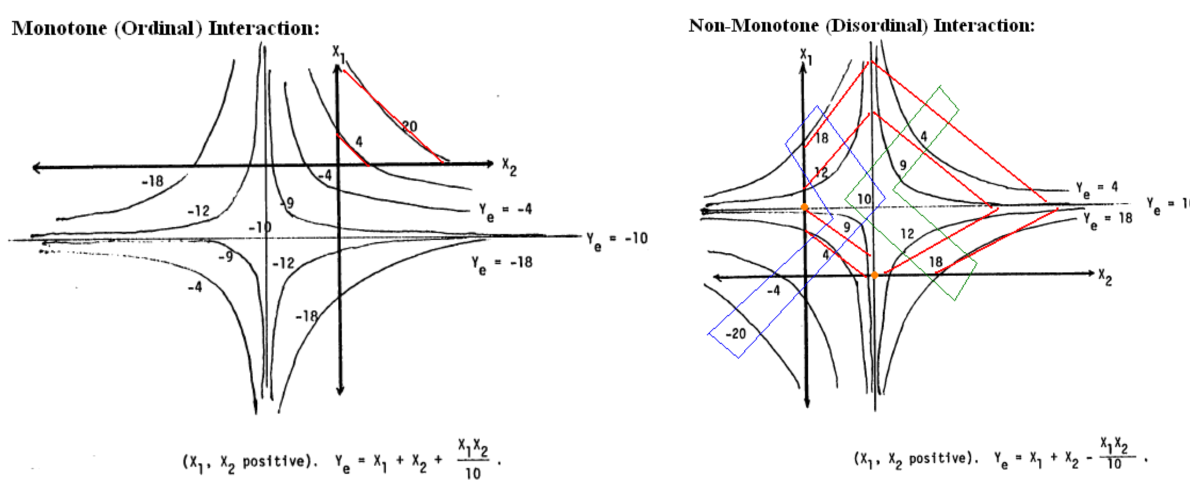


Figure 5. Linear Modeling of Monotone Versus Non-Monotone Relationships (Example from Hastie & Dawes, 2001)



(reproduced with modification from Hastie & Dawes, 2001, Fig. 3.2, p. 59)

Figure 6. Linear Modeling of Monotone Versus Non-Monotone Relationships (Example from Camerer, 1981a)



(reproduced with modification from Camerer, 1981a, Figs. 3-4, pp. 27-28)

Figure 7. Optimal Nonlinear Function Becoming More Linear With Increasing Error in Measurement of the Independent Variables

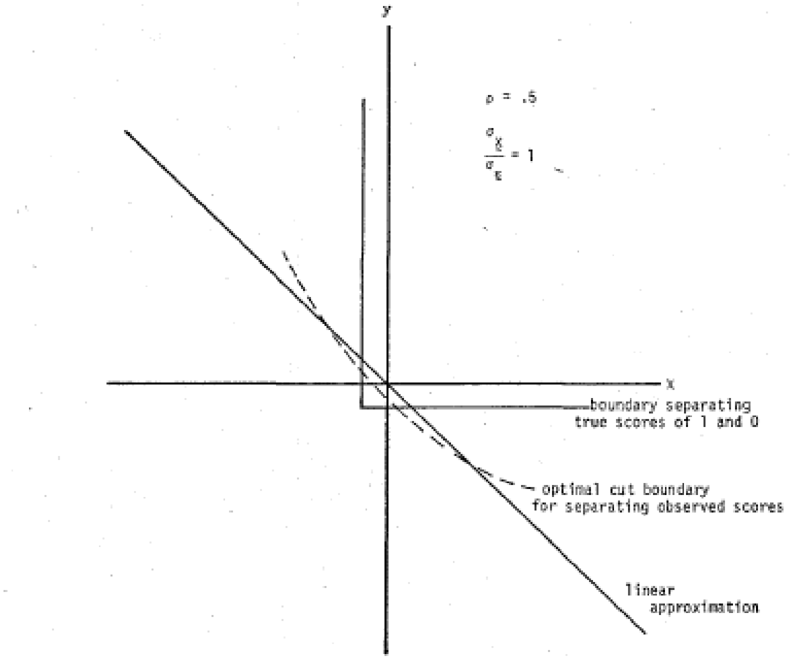
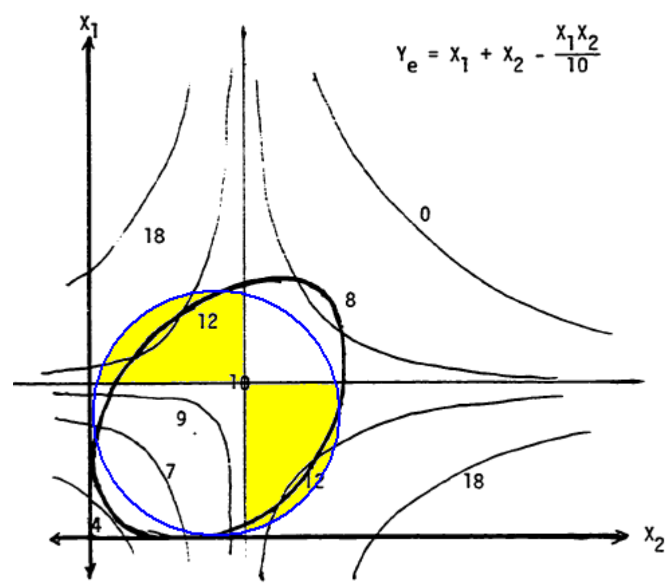


FIGURE 1. Conjunctive true score region, optimal cut boundary, and linear approximation.

(reproduced from Dawes and Corrigan, 1974)

Figure 8. Positive Cue Redundancy Attenuating the Size of the Disordinal Interaction



Positive redundancy
Versus Orthogonality

(reproduced with modification from Camerer, 1981a, Fig. 5, p. 29)

Figure 9. Disordinal Interaction in the Fall 2009 and Spring 2010 Experiments

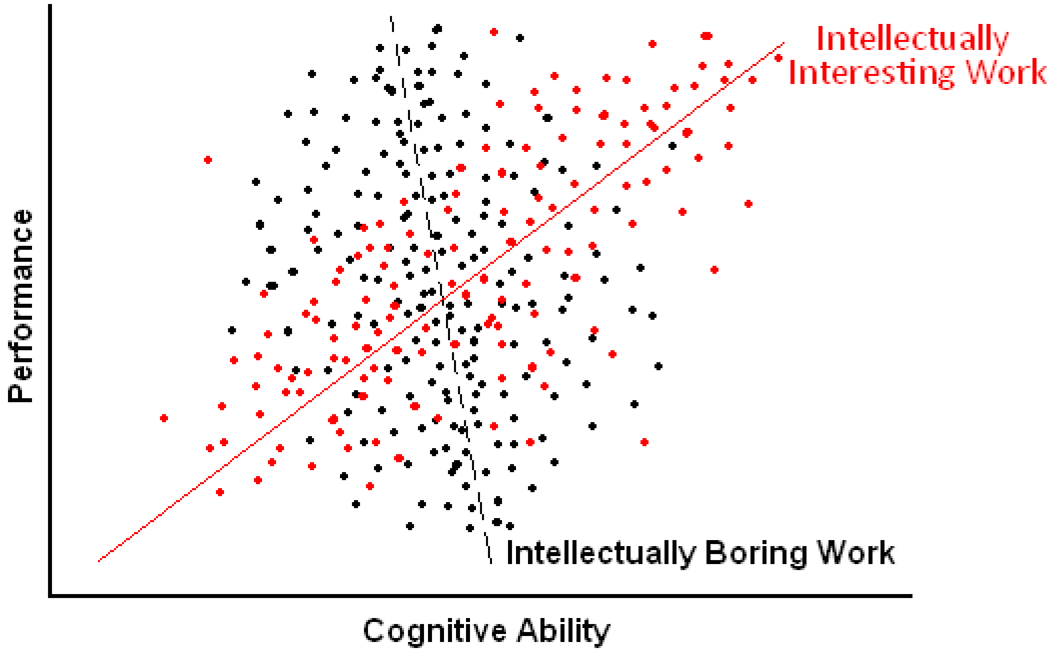


Figure 10. Prediction Accuracy After Training: r_a (Fall 2009)

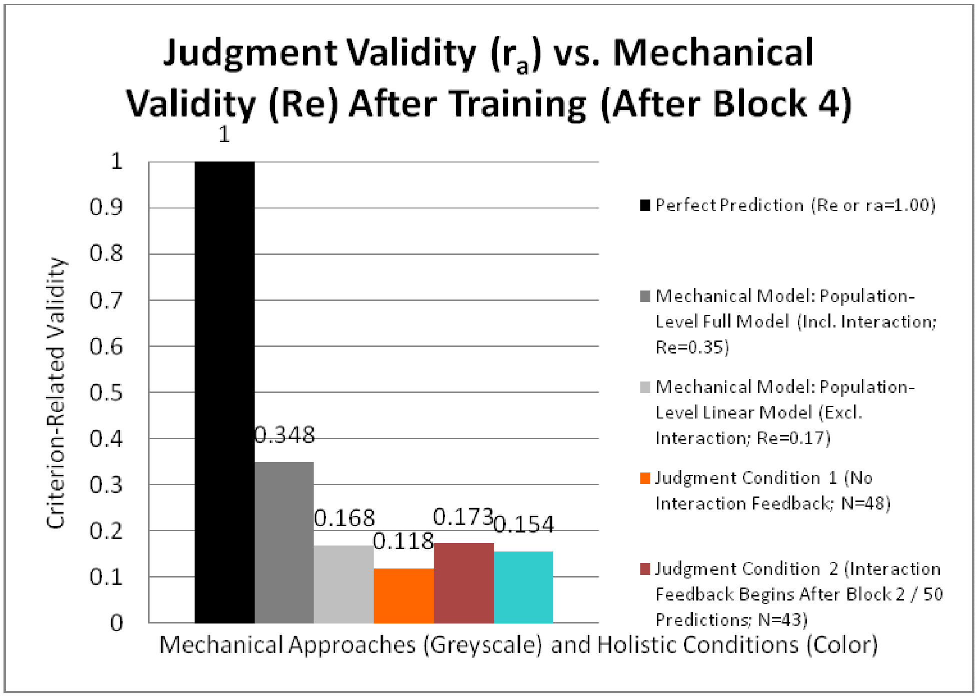


Figure 11. Prediction Accuracy After Training: Skill Score (Fall 2009)

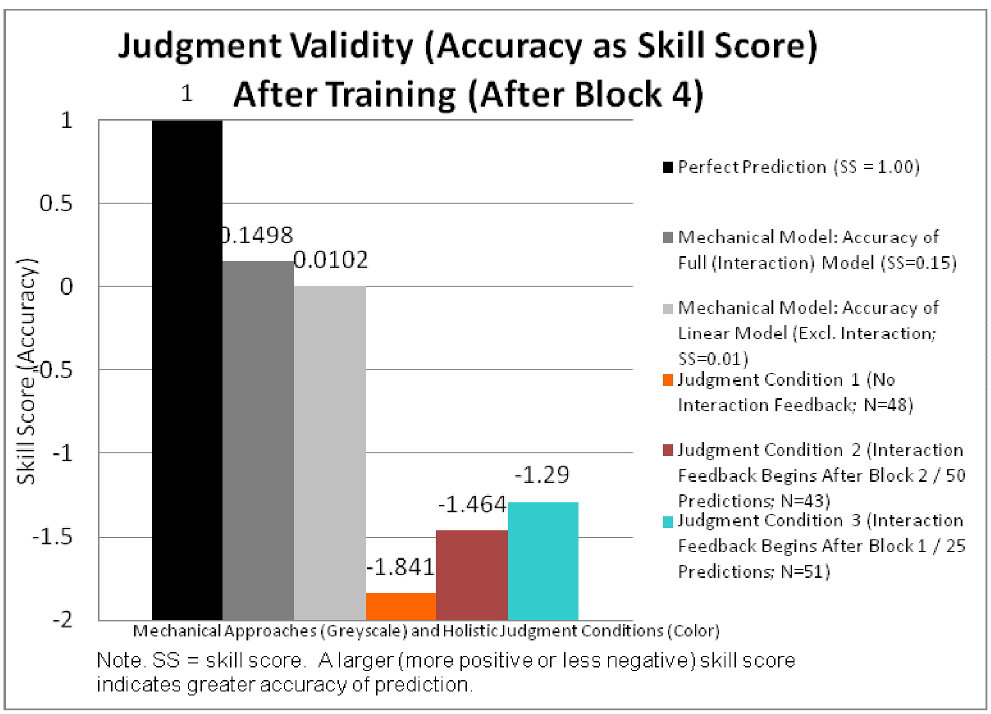


Figure 12. Prediction Accuracy During Training: r_a (Fall 2009)

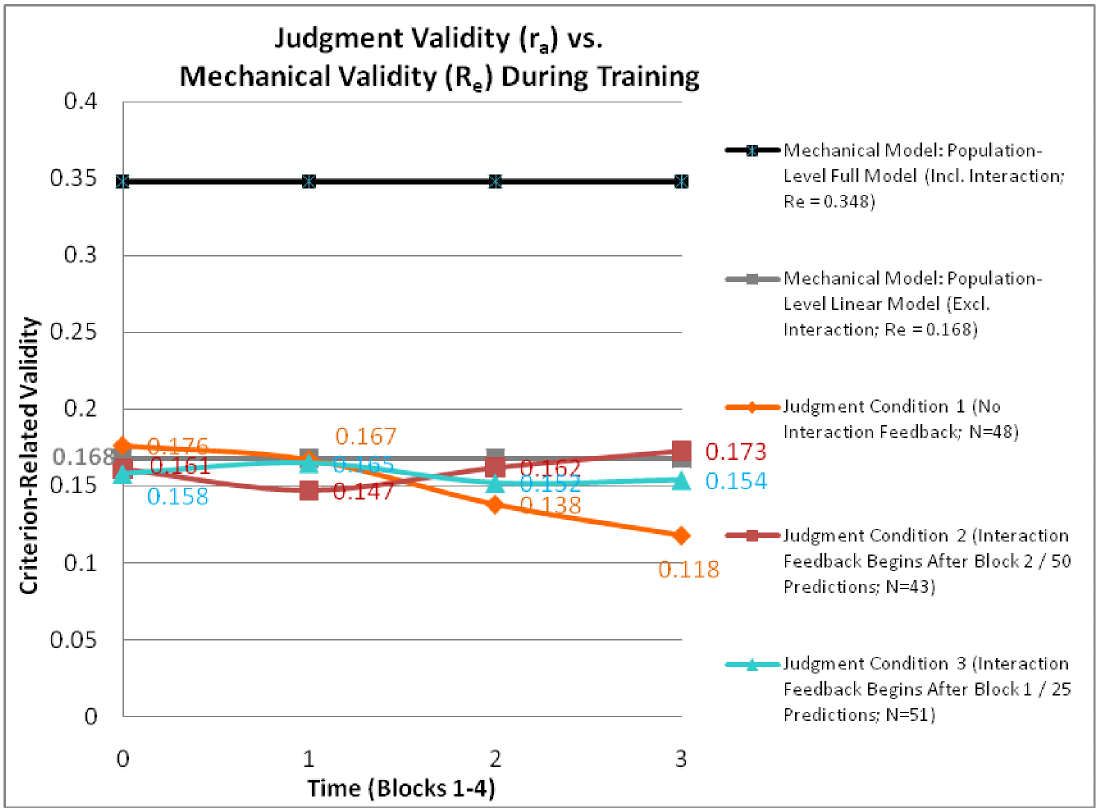


Figure 13. Prediction Accuracy During Training: Skill Score (Fall 2009)

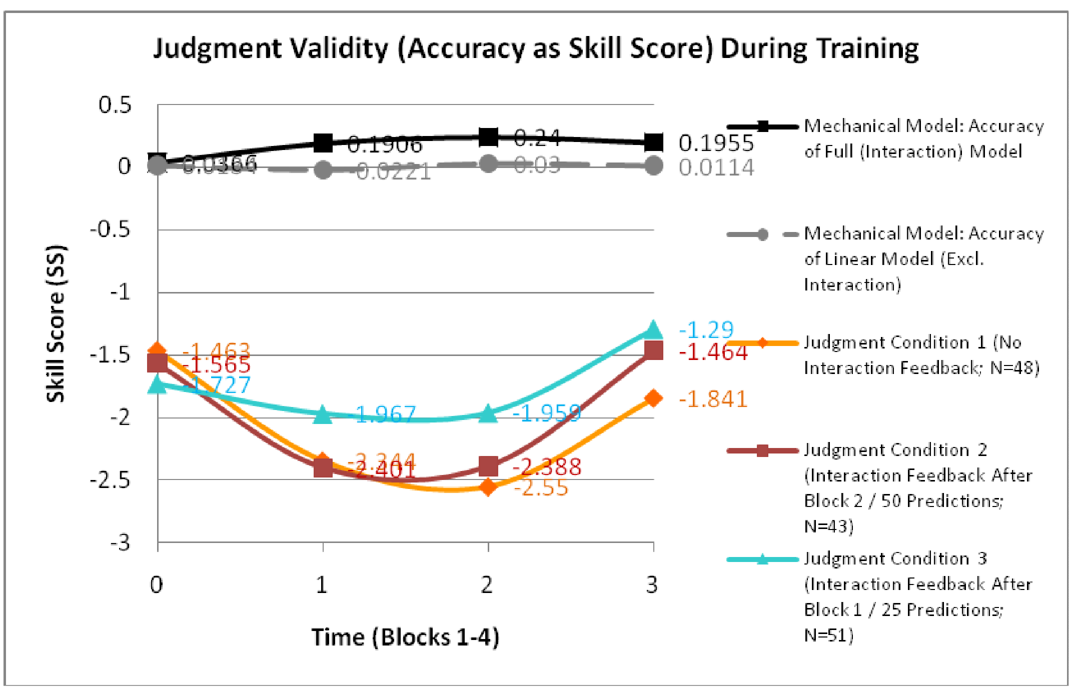


Figure 14. Change Over Time in Fisher r_a (Judgment Criterion-Related Validity): Trellis Plot of Slopes and Intercepts (*N* = All 142 Subjects) (Fall 2009)

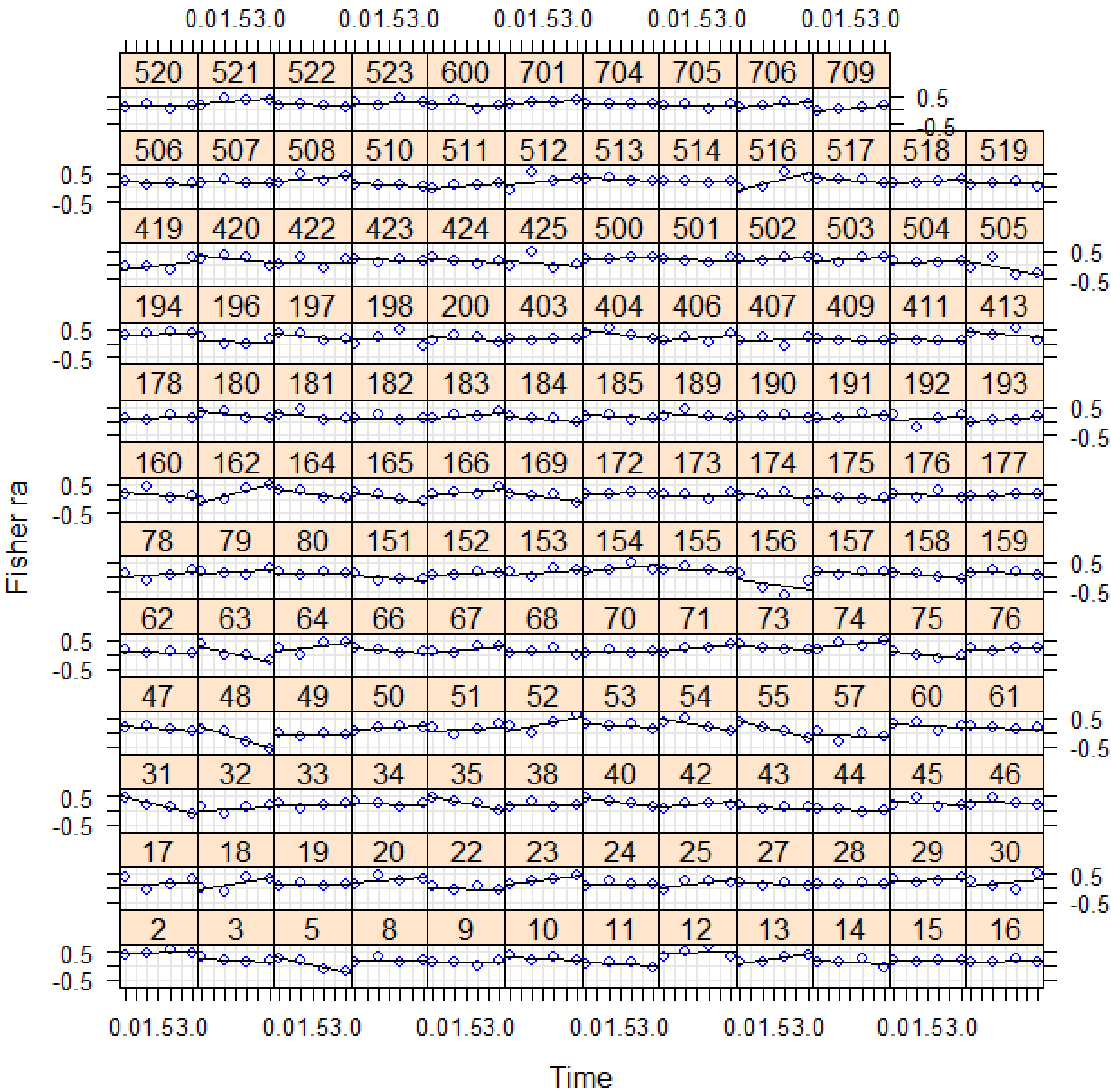


Figure 15. Change Over Time in Fisher r_a (Judgment Criterion-Related Validity): Boxplots of Slopes and Intercepts for Feedback Conditions (C1, C2, and C3) (Fall 2009)

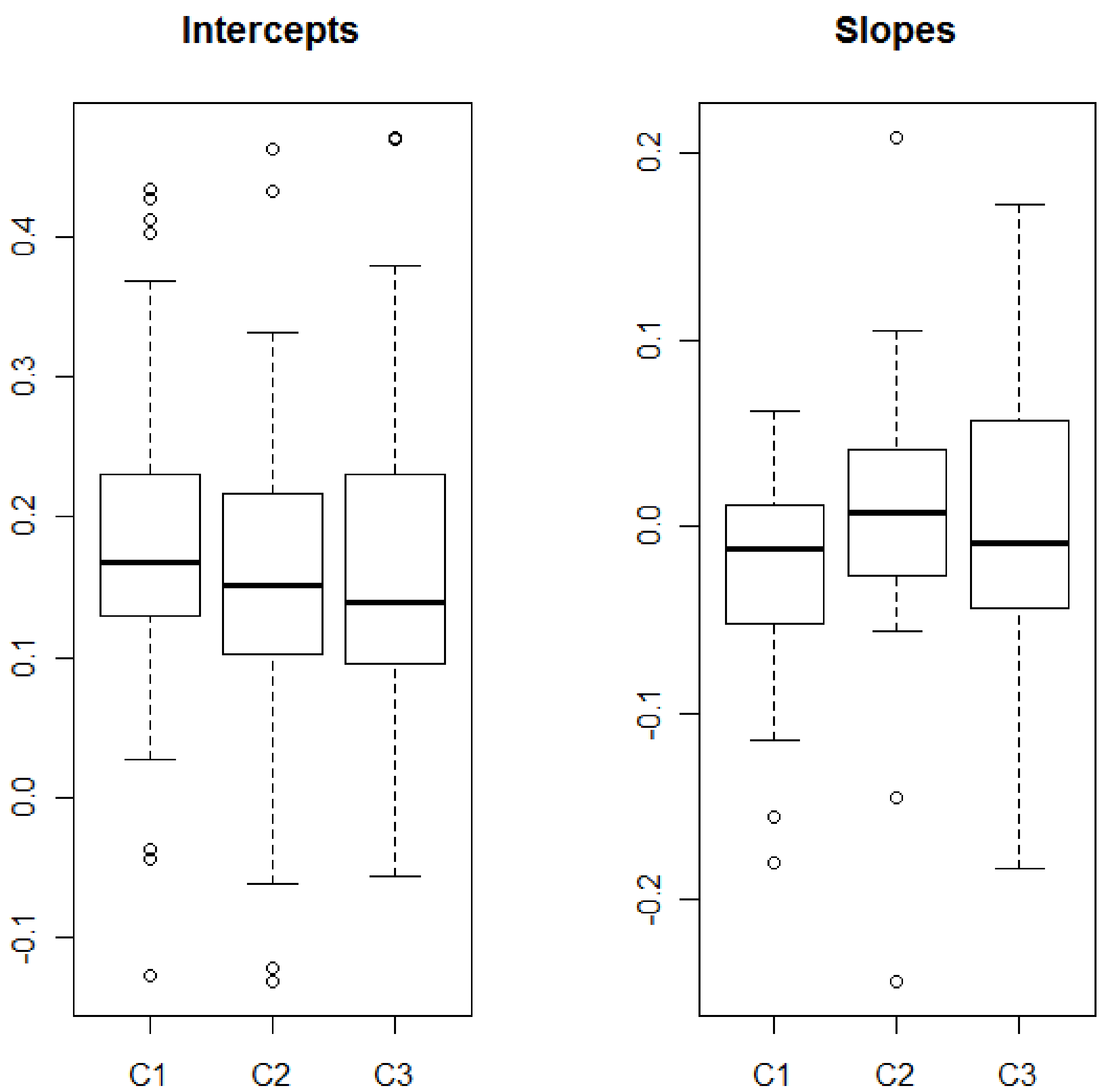


Figure 16. Change Over Time in Fisher r_a (Judgment Criterion-Related Validity): Subjects' Individual Regression Lines (N = All 142 Subjects) (Fall 2009)

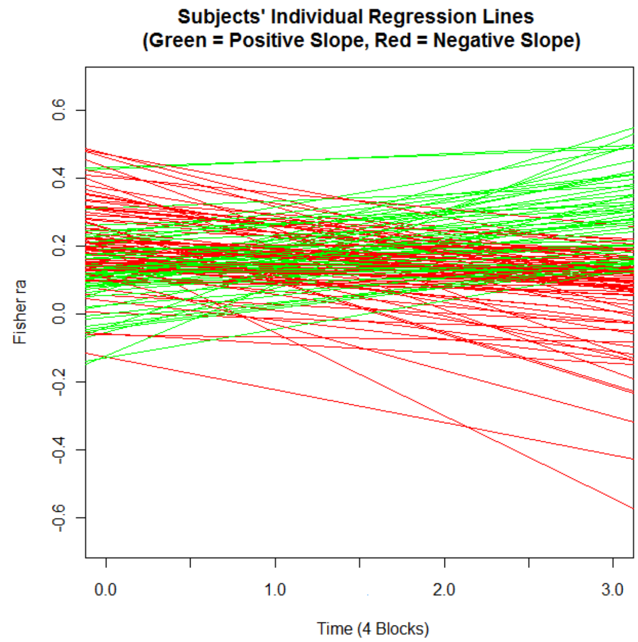


Figure 17. Change Over Time in Fisher r_a (Judgment Criterion-Related Validity): Regression Line For Fixed Effect (Random Effects Ignored; No Interaction) (N = All 142 Subjects) (Fall 2009)

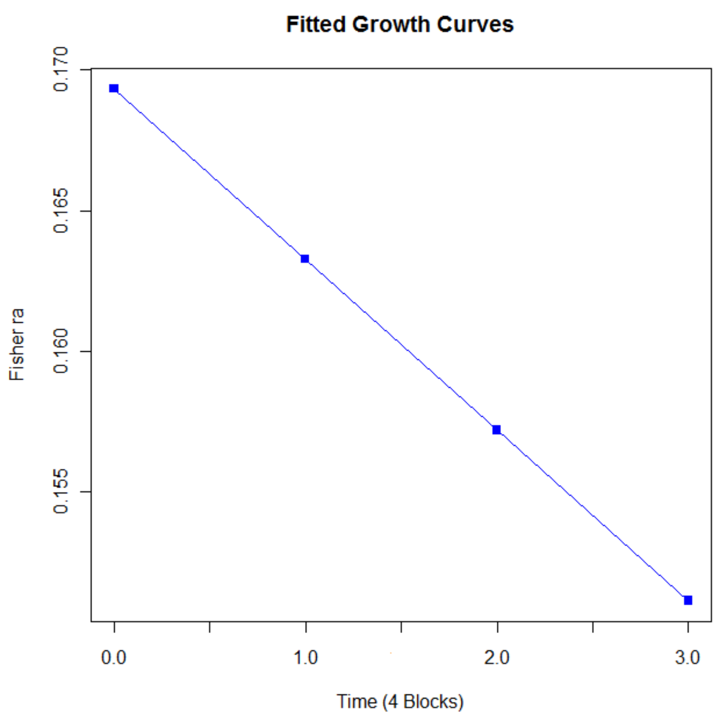


Figure 18. Change Over Time in Skill Score (Accuracy): Trellis Plot of Slopes and Intercepts (Fall 2009) (N = All 142 Subjects)

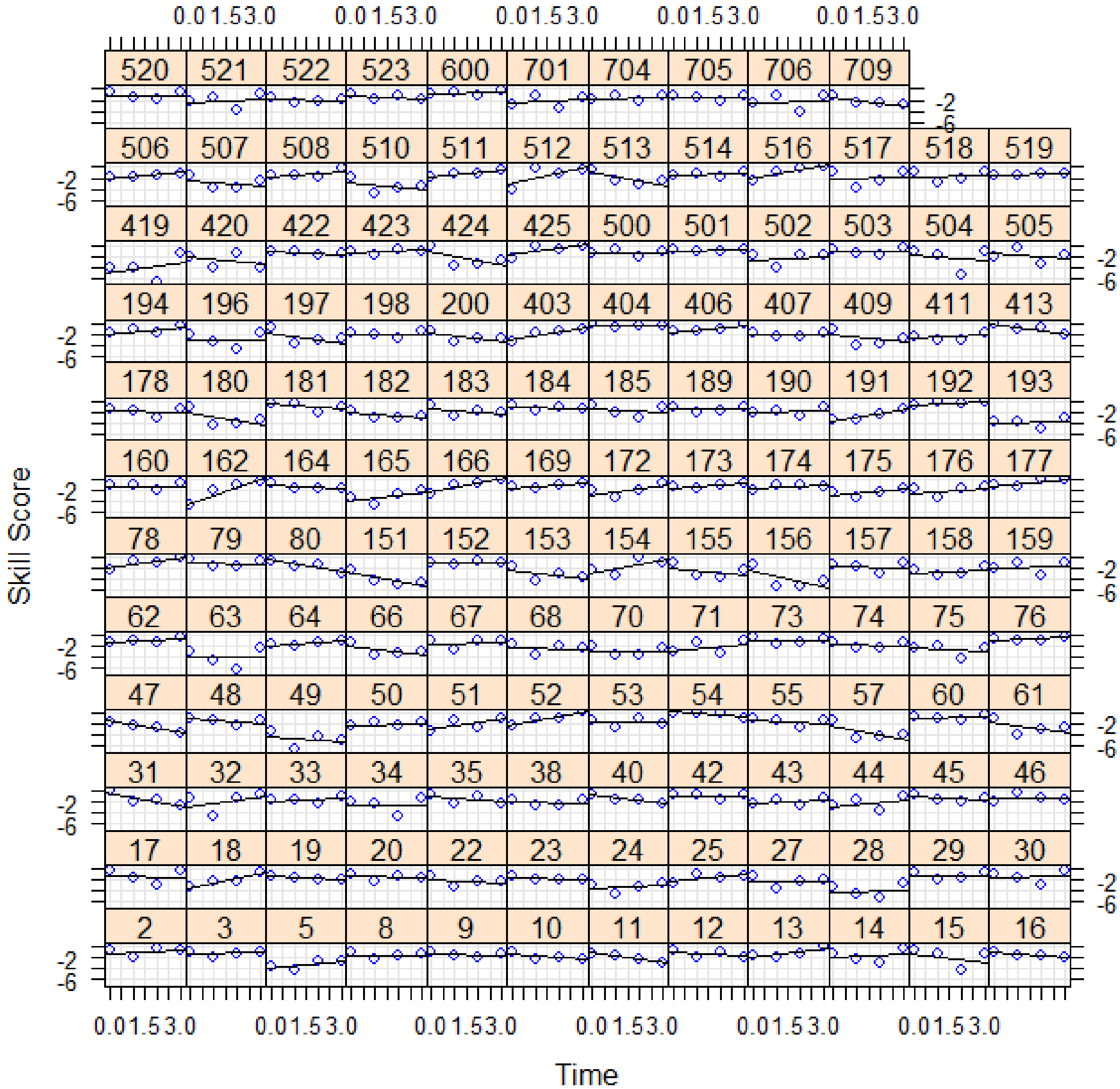


Figure 19. Change Over Time in Skill Score (Accuracy): Boxplots of Slopes and Intercepts for Feedback Conditions (C1, C2, and C3) (N = All 142 Subjects) (Fall 2009)

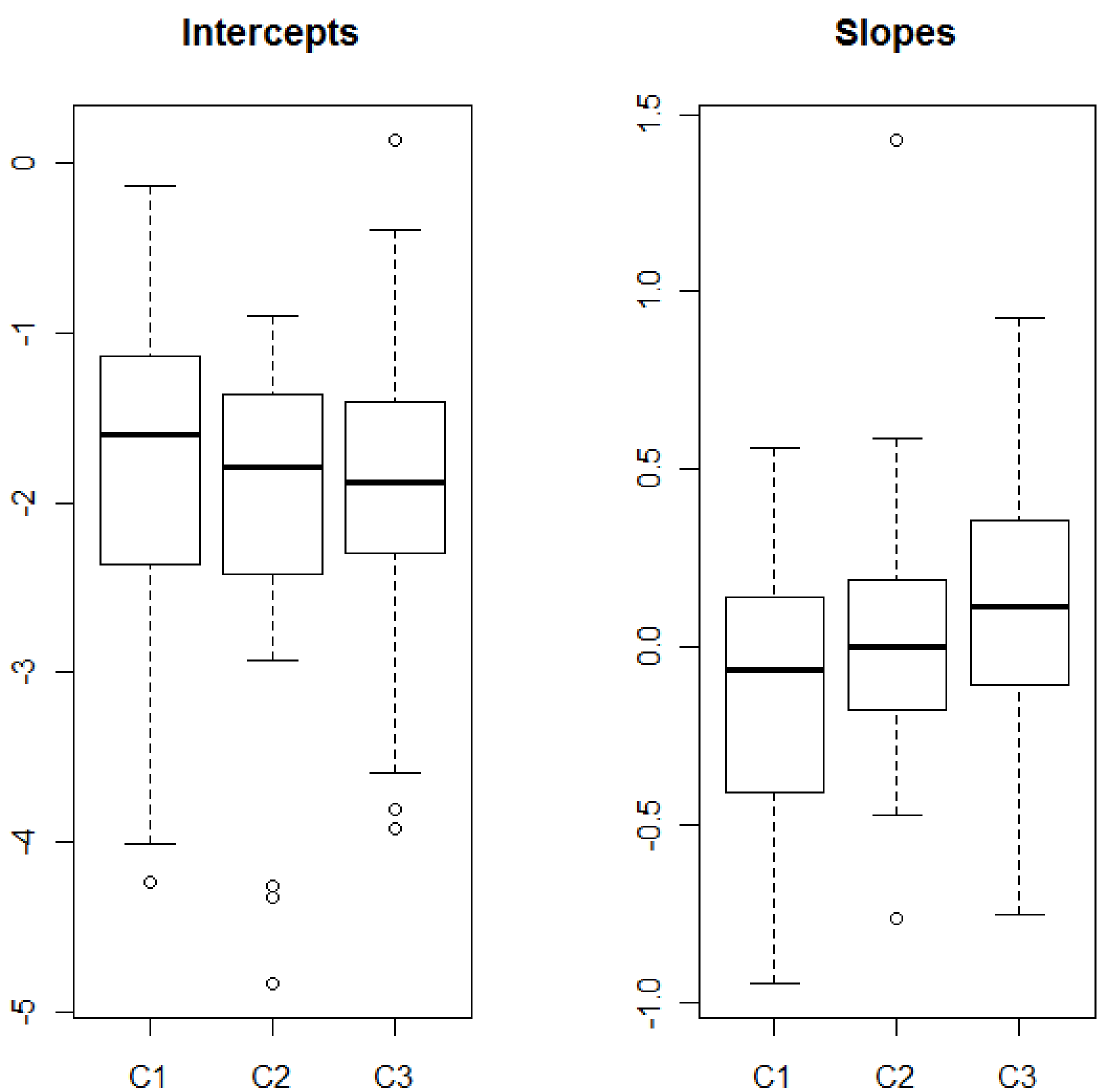


Figure 20. Change Over Time in Skill Score (Accuracy): Fitted Growth Curves (Fall 2009)

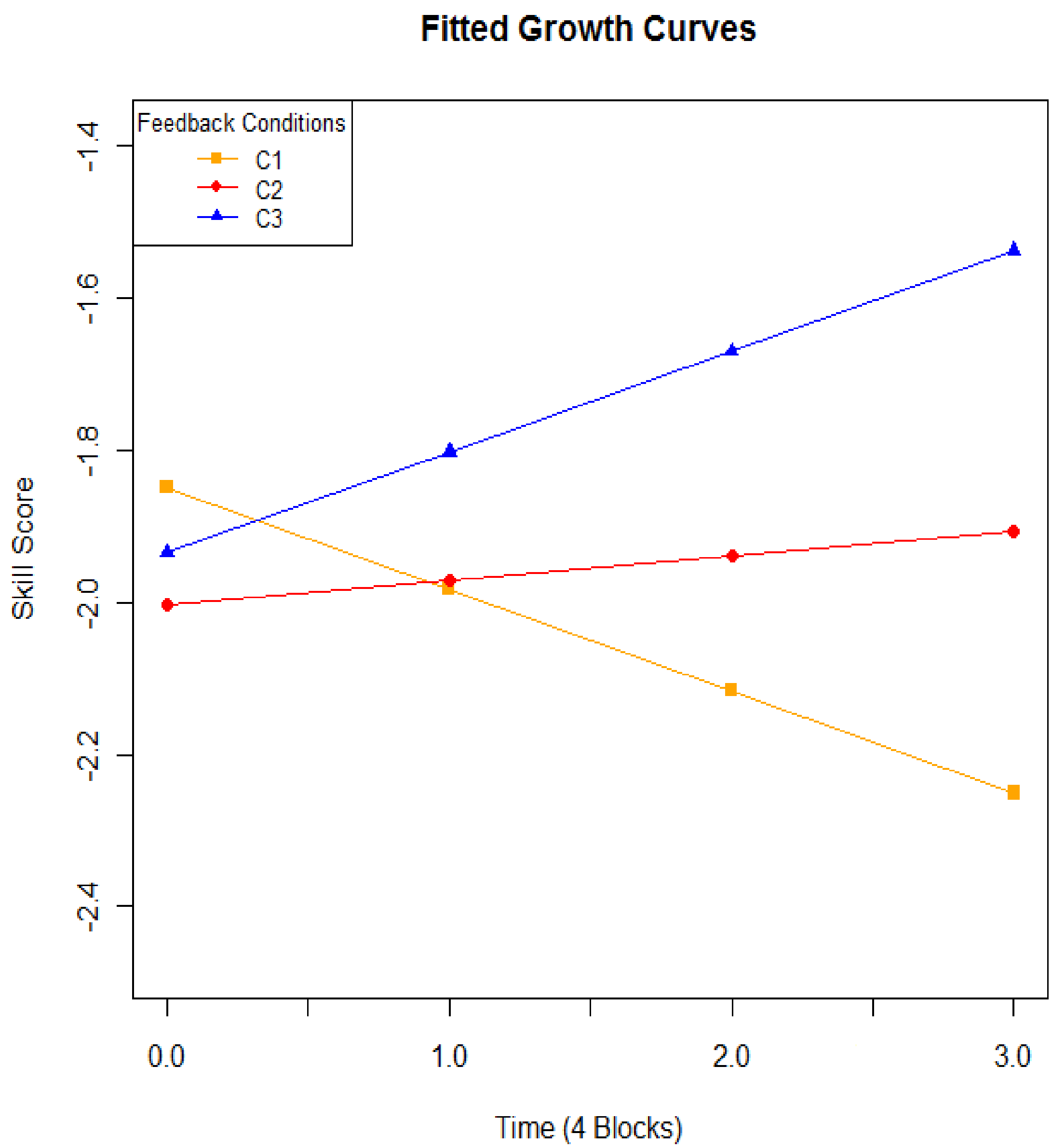


Figure 21. Change Over Time in Fisher *C* (Unmodeled Knowledge): Trellis Plot of Slopes and Intercepts (*N* = All 142 Subjects) (Fall 2009)

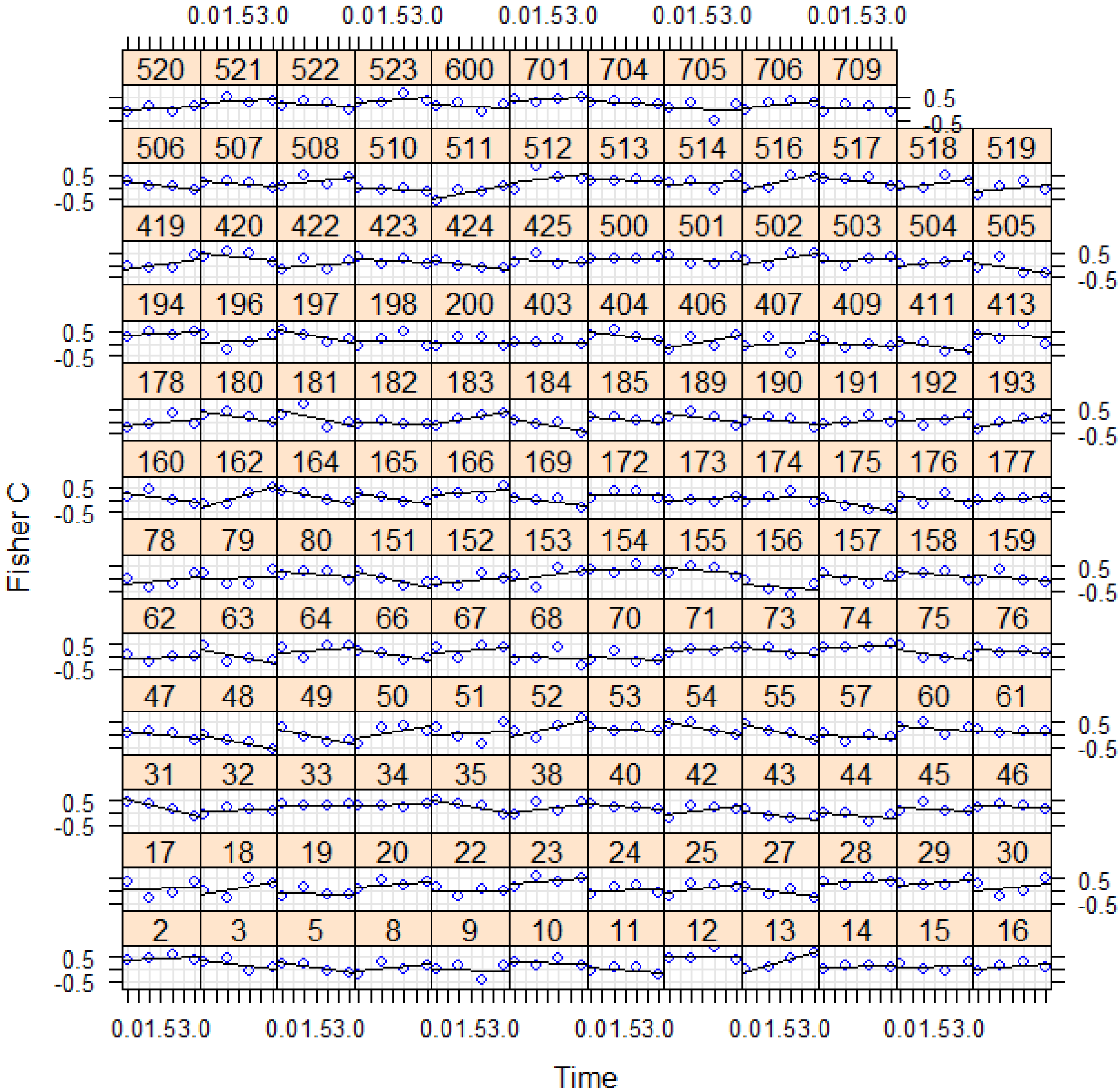


Figure 22. Change Over Time in Fisher *C* (Unmodeled Knowledge): Boxplots of Slopes and Intercepts for Feedback Conditions (C1, C2, and C3) (Fall 2009)

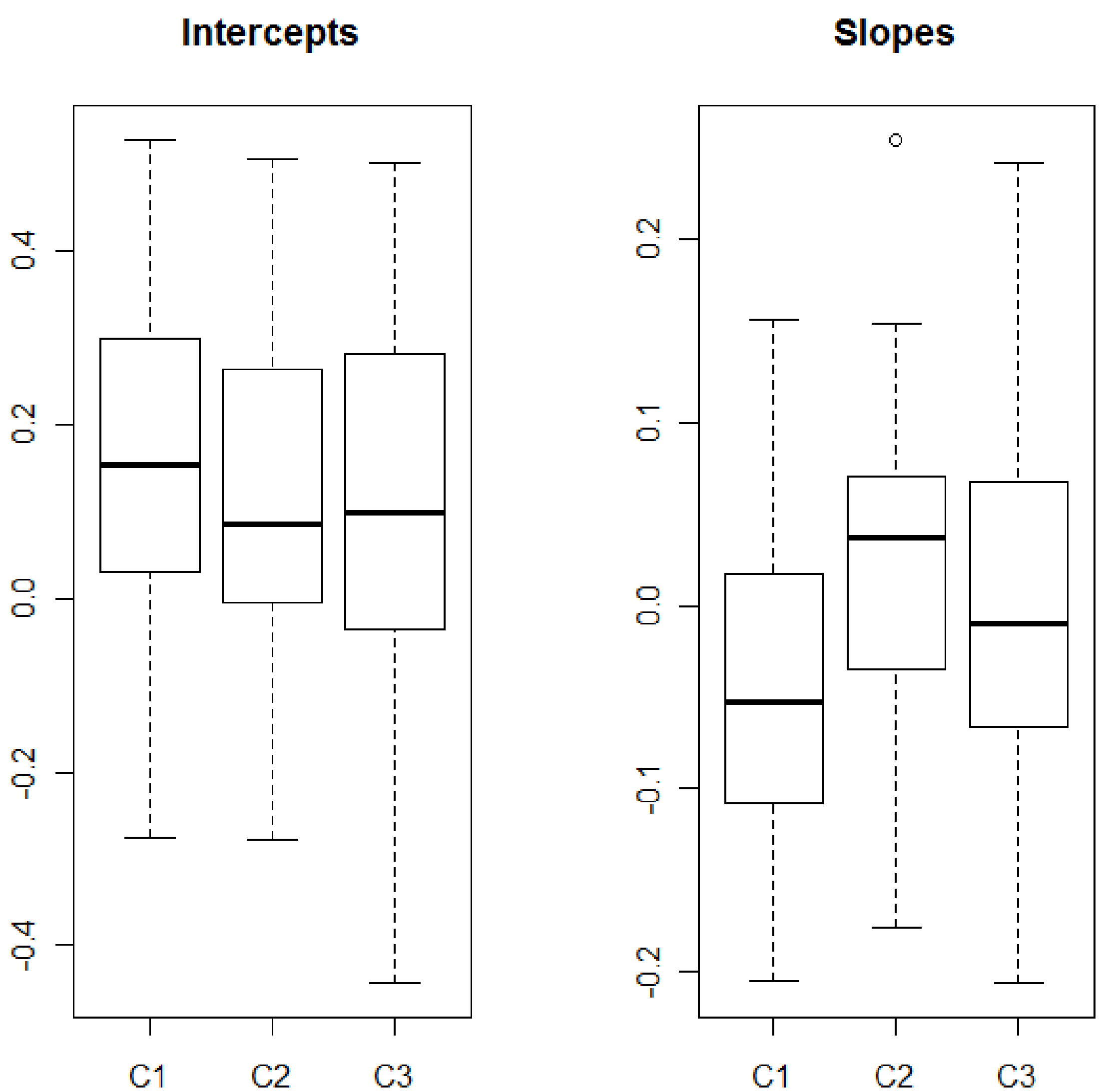


Figure 23. Change Over Time in Fisher *C* (Unmodeled Knowledge): Fitted Growth Curves (Fall 2009)

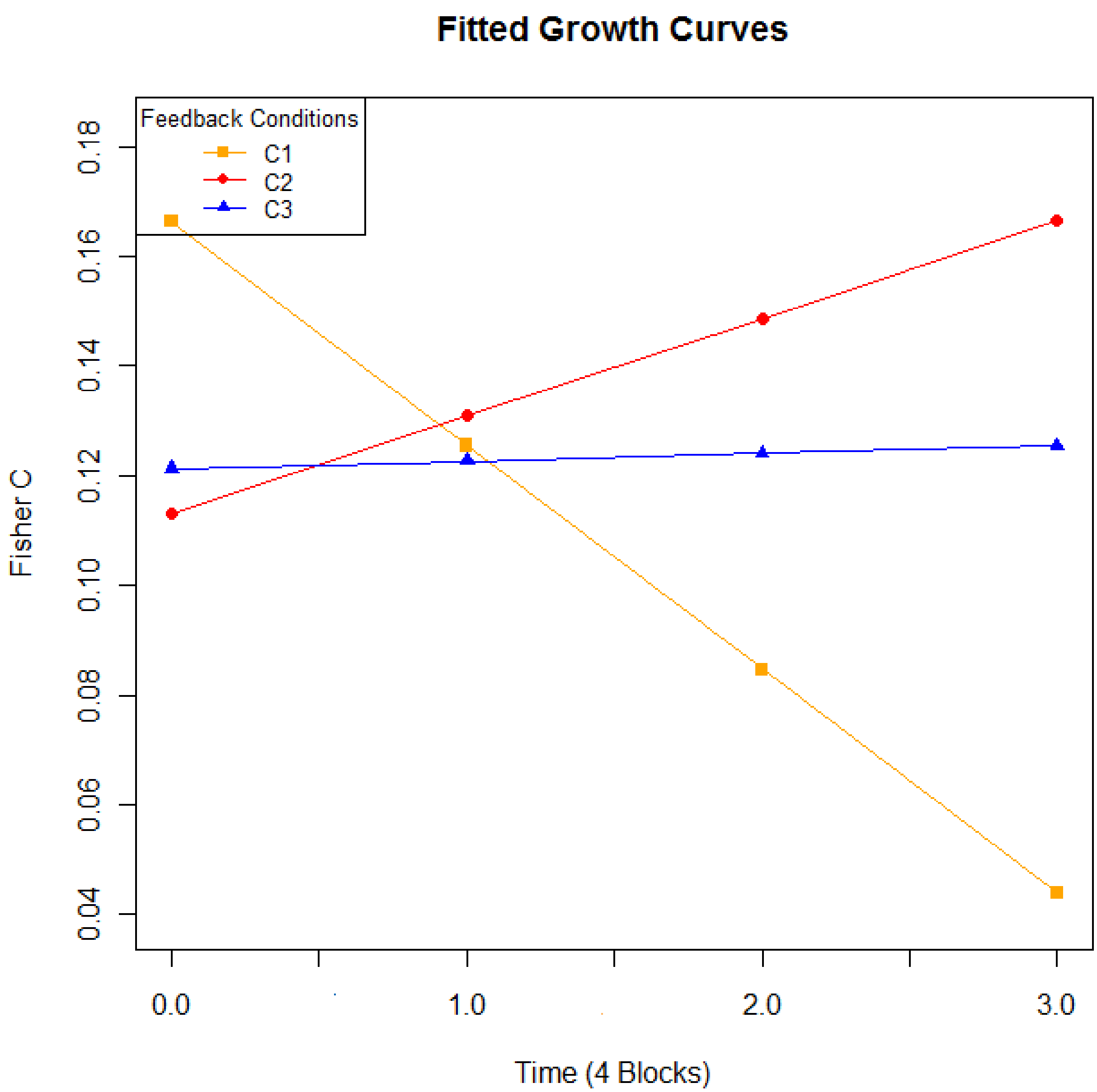


Figure 24. Change Over Time in Fisher r_z (Criterion-Related Validity of Unmodeled Knowledge): Trellis Plot of Slopes and Intercepts (N = All 142 Subjects) (Fall 2009)

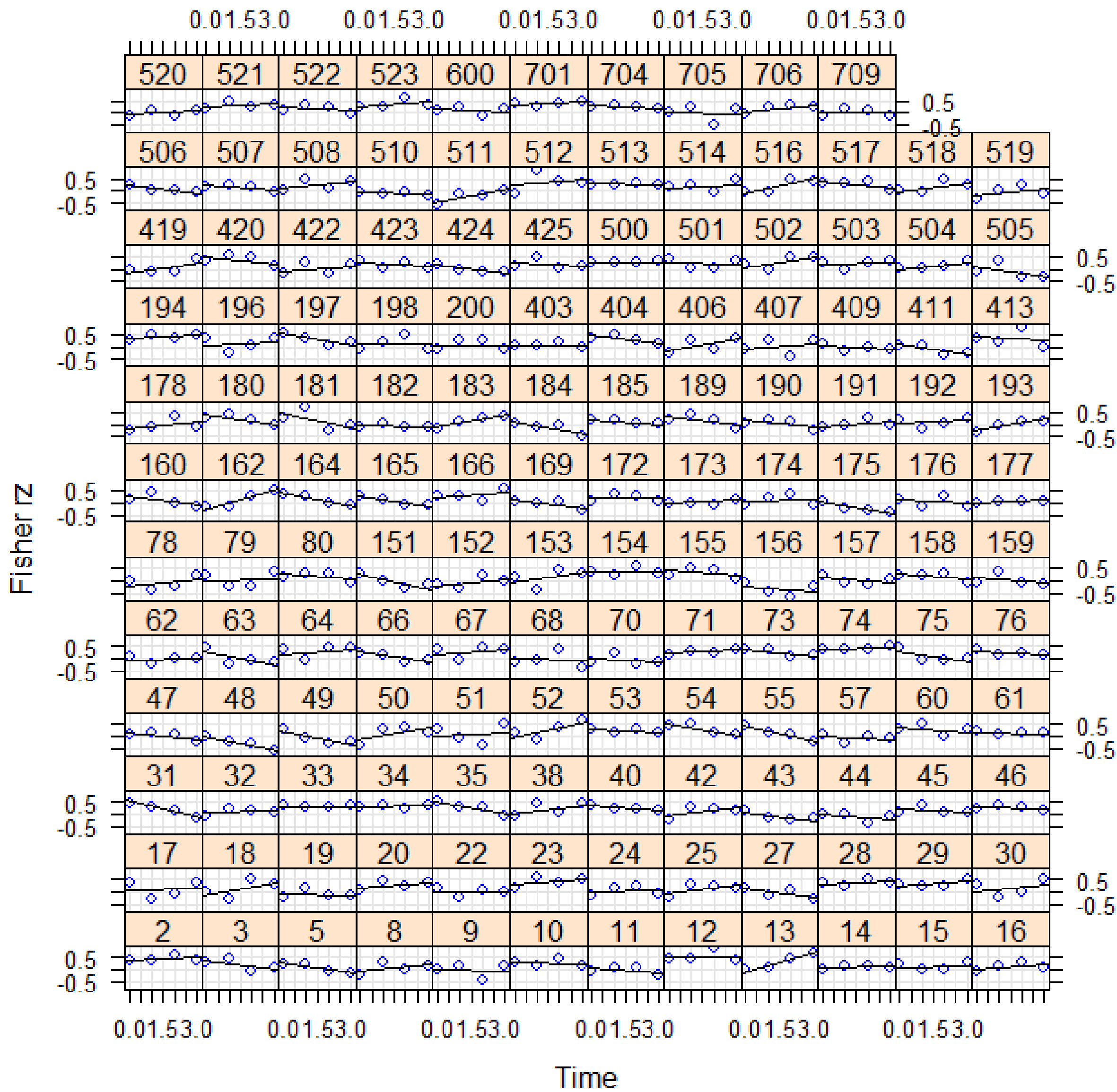


Figure 25. Change Over Time in Fisher r_z (Criterion-Related Validity of Unmodeled Knowledge): Boxplots of Slopes and Intercepts for Feedback Conditions (C1, C2, and C3) (Fall 2009)

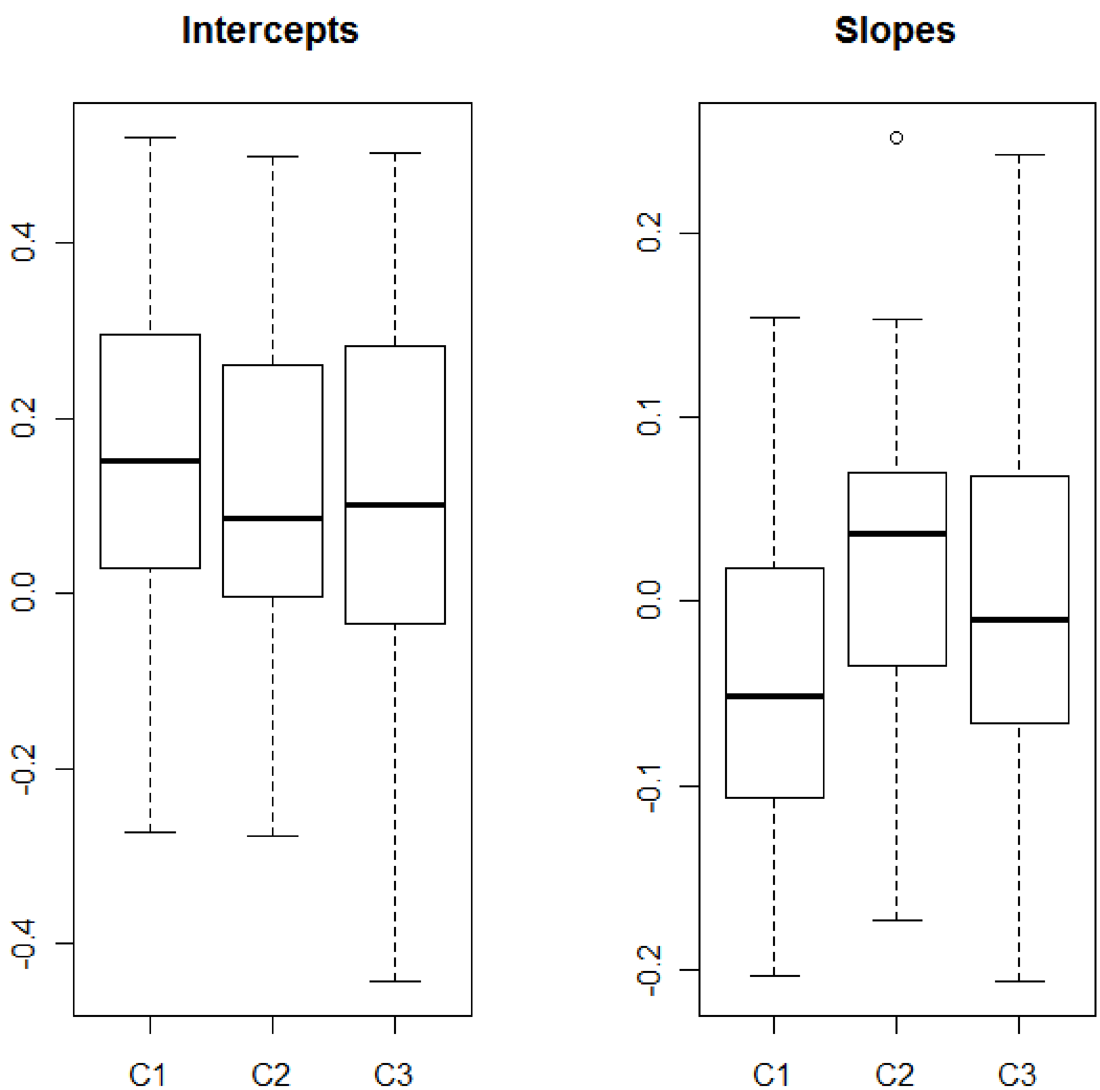


Figure 26. Change Over Time in Fisher r_z (Criterion-Related Validity of Unmodeled Knowledge): Fitted Growth Curves (Fall 2009)

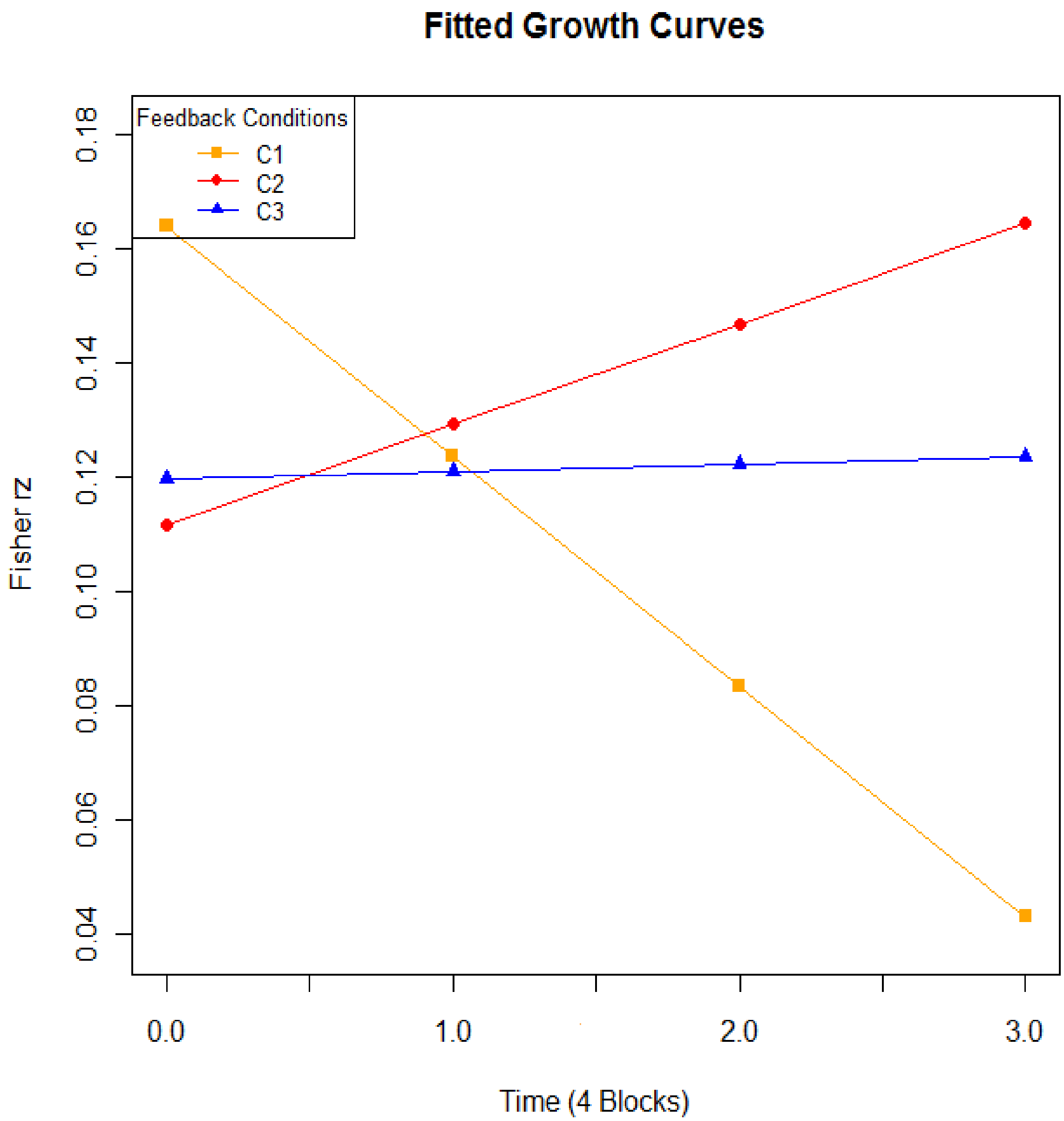


Figure 27. Change Over Time in Fisher *G* (Mechanical Knowledge): Trellis Plot of Slopes and Intercepts (*N* = All 142 Subjects) (Fall 2009)

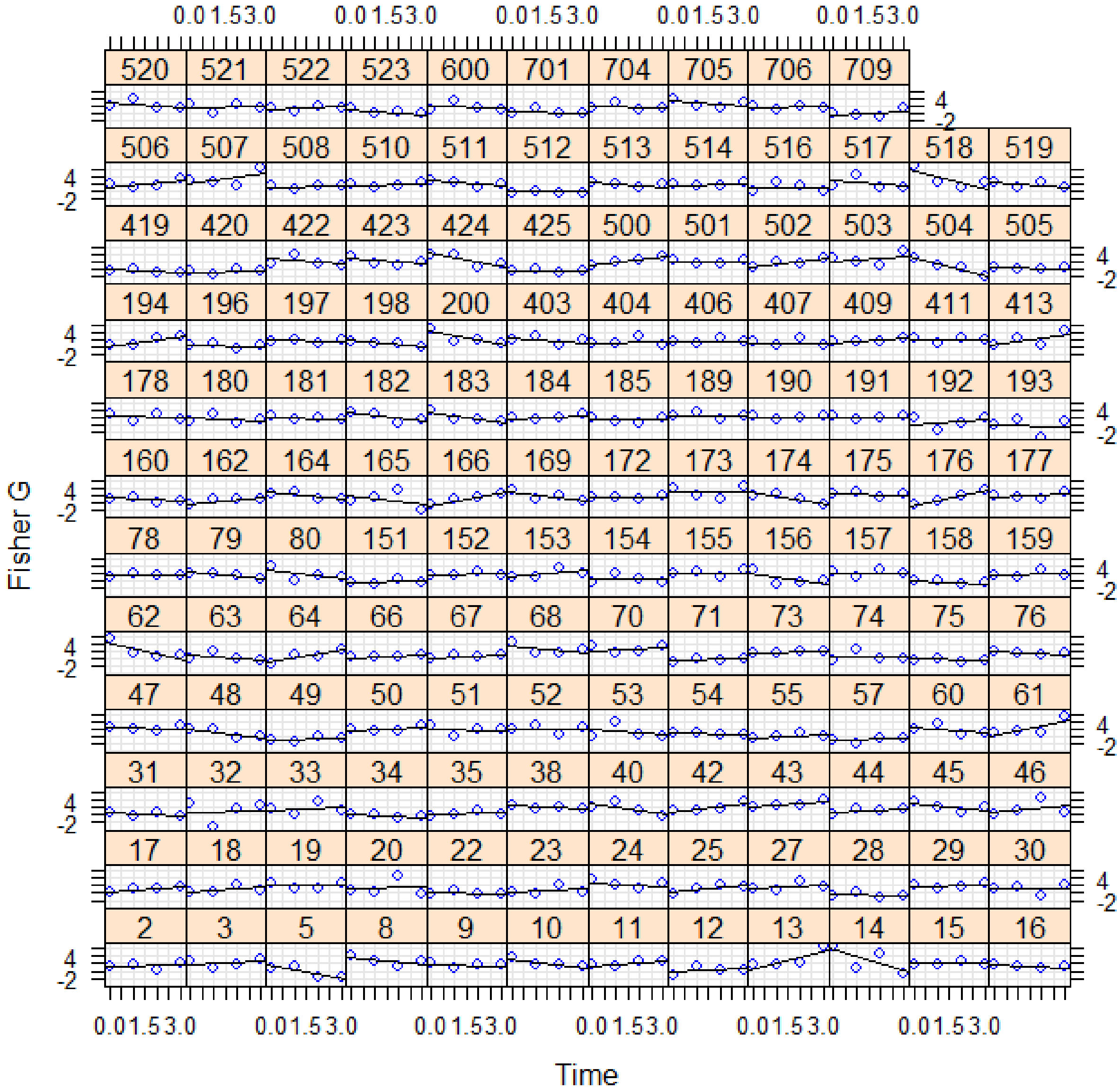


Figure 28. Change Over Time in Fisher *G* (Mechanical Knowledge): Boxplots of Slopes and Intercepts for Feedback Conditions (C1, C2, and C3) (Fall 2009)

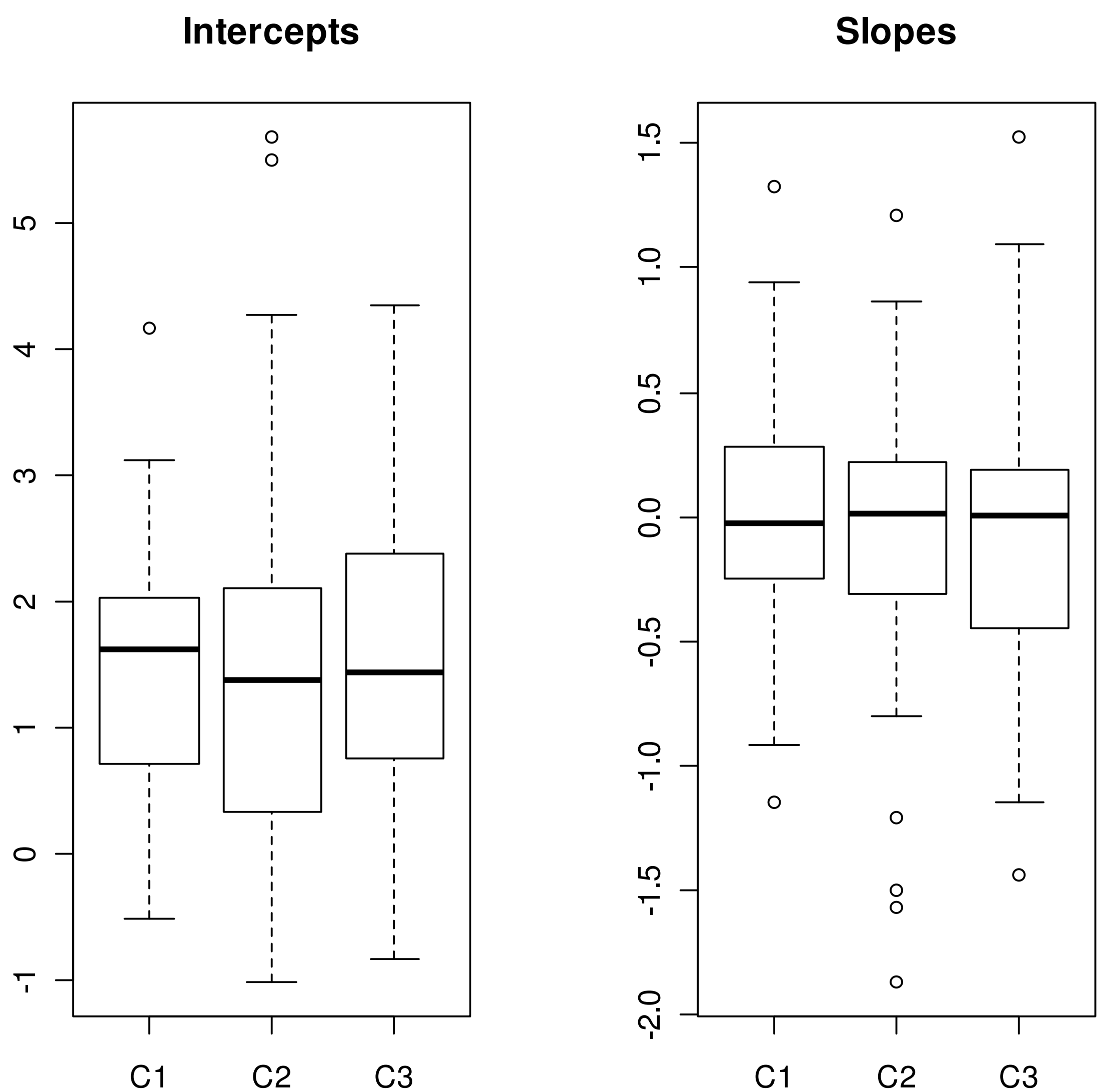


Figure 29. Change Over Time in Fisher *G* (Mechanical Knowledge): Fitted Growth Curve (Fixed Effect Only; No Random Effects; No Interaction) (*N* = All 142 Subjects) (Fall 2009)

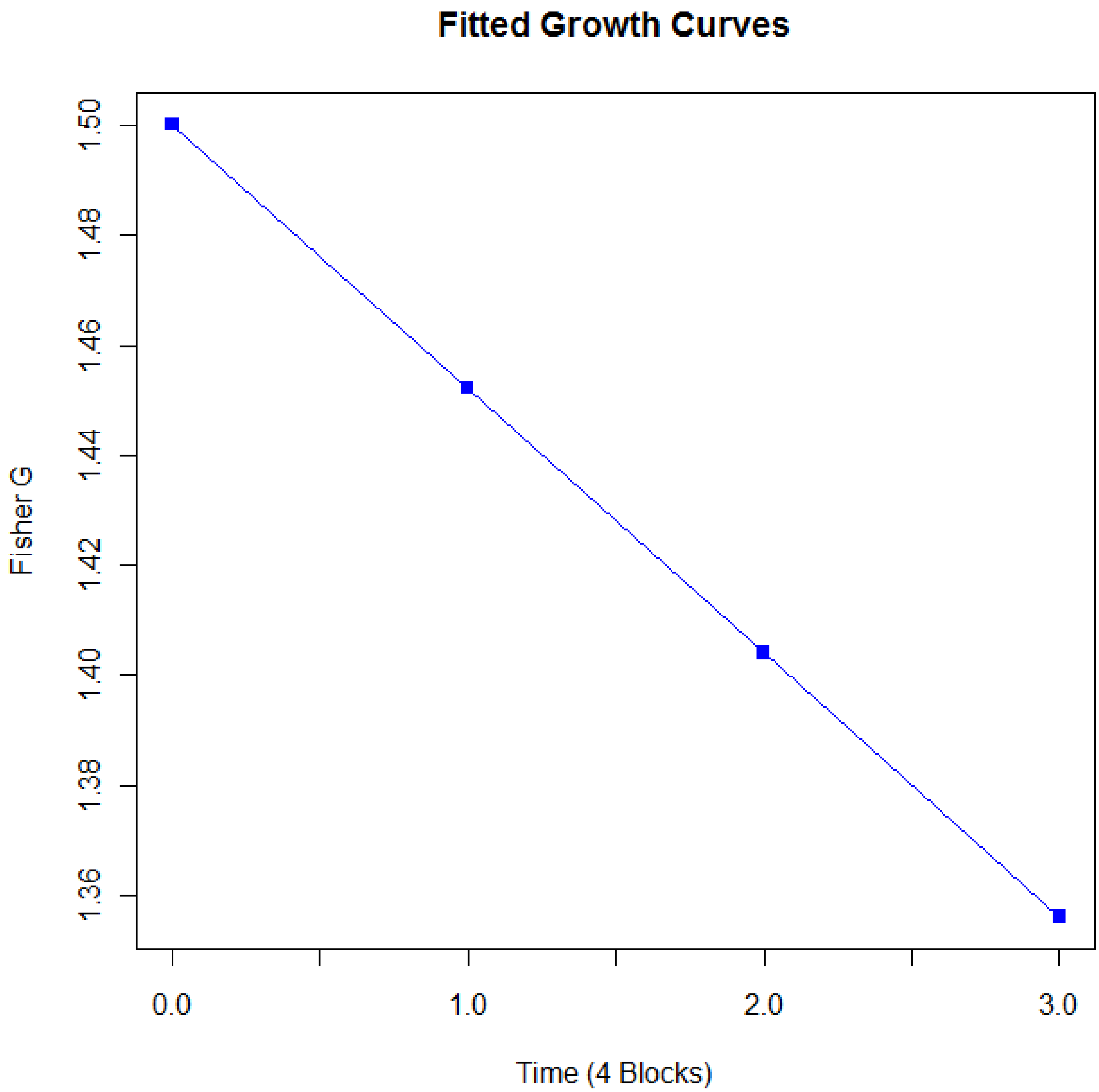


Figure 30. Change Over Time in Fisher R_s (Cognitive Control): Trellis Plot of Slopes and Intercepts (N = All 142 Subjects) (Fall 2009)

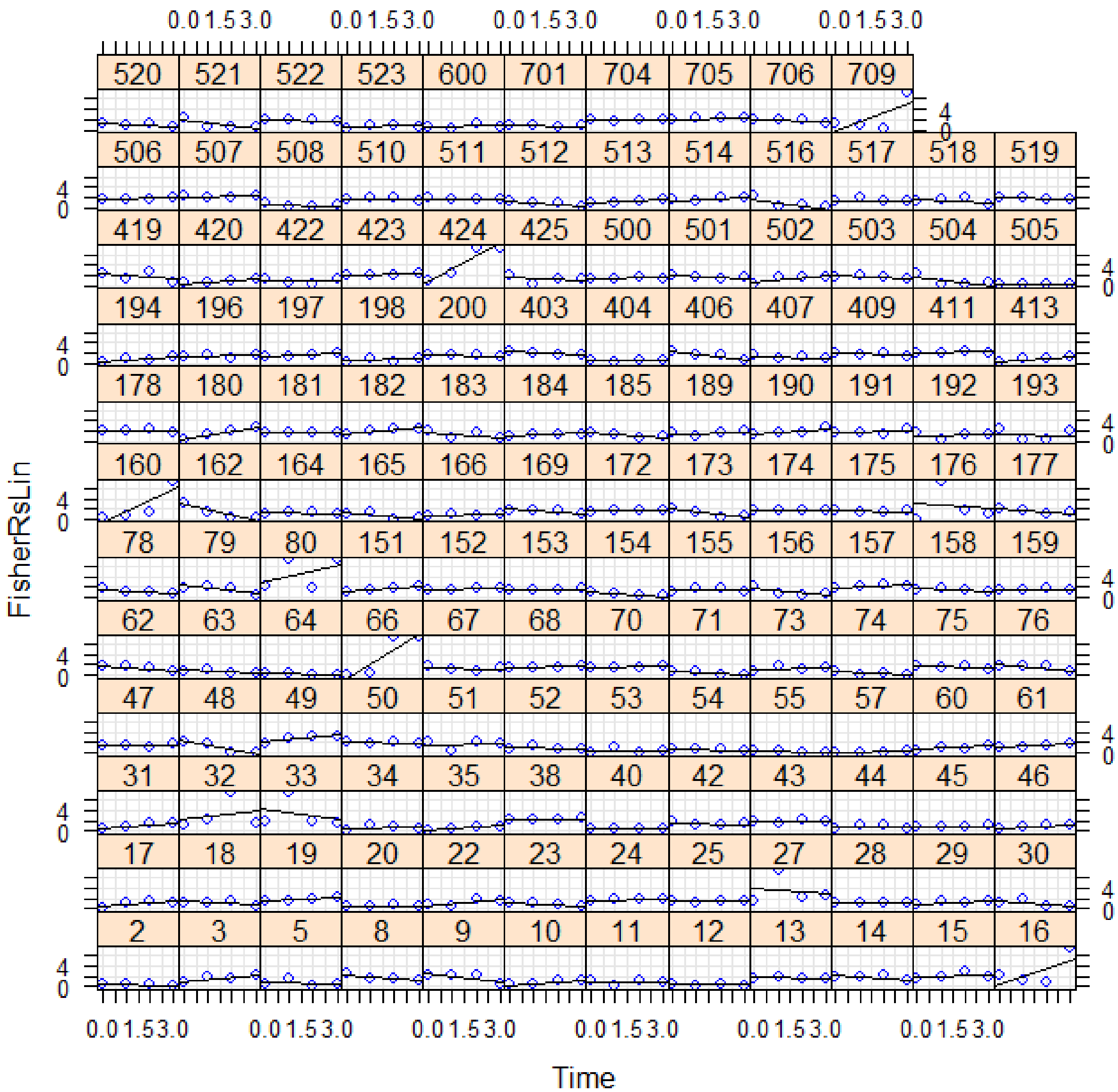


Figure 31. Change Over Time in Fisher R_s (Cognitive Control): Boxplots of Slopes and Intercepts for Feedback Conditions (C1, C2, and C3) (Fall 2009)

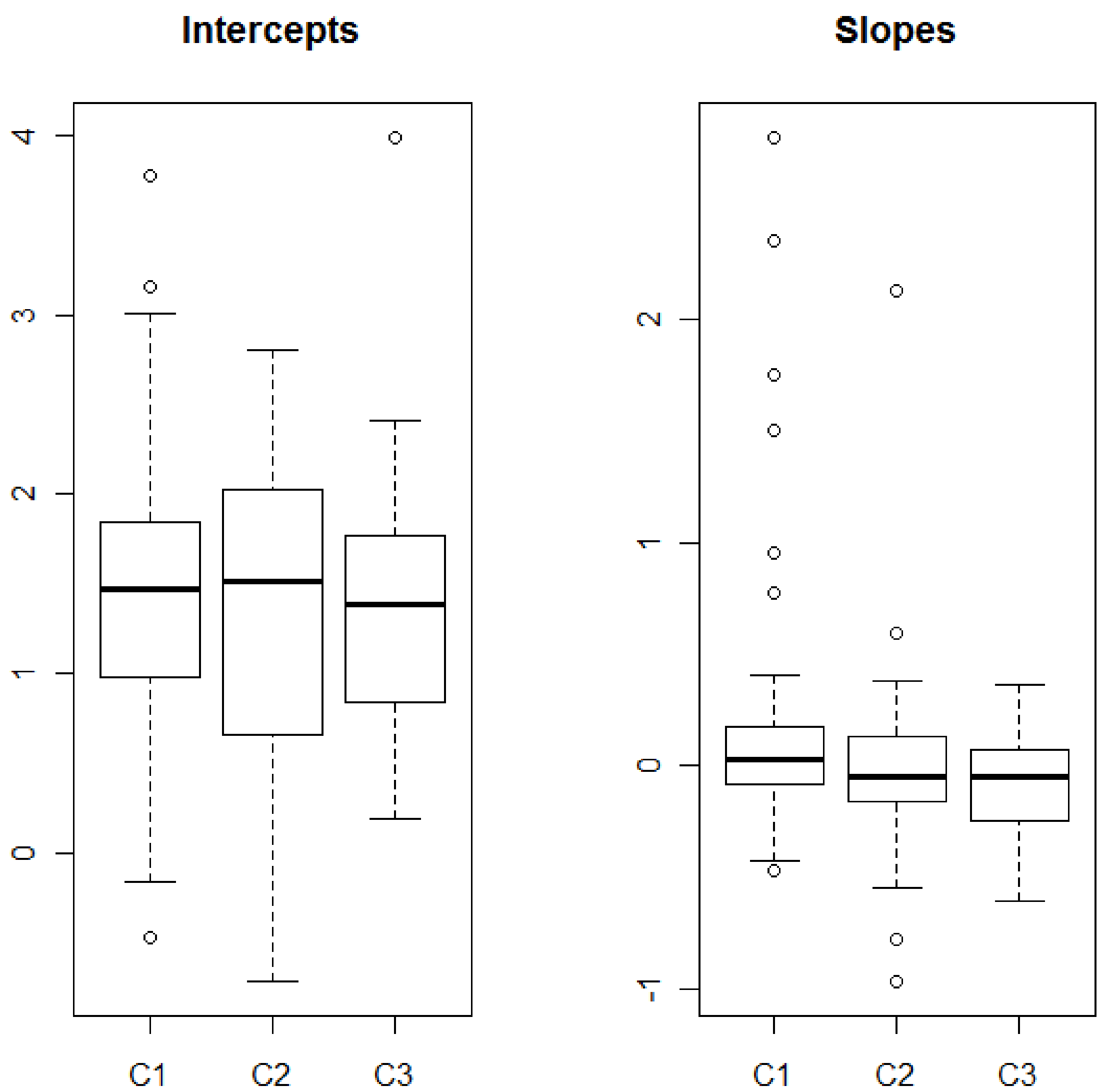


Figure 32. Change Over Time in Fisher R_s (Cognitive Control): Fitted Growth Curves (Fall 2009)

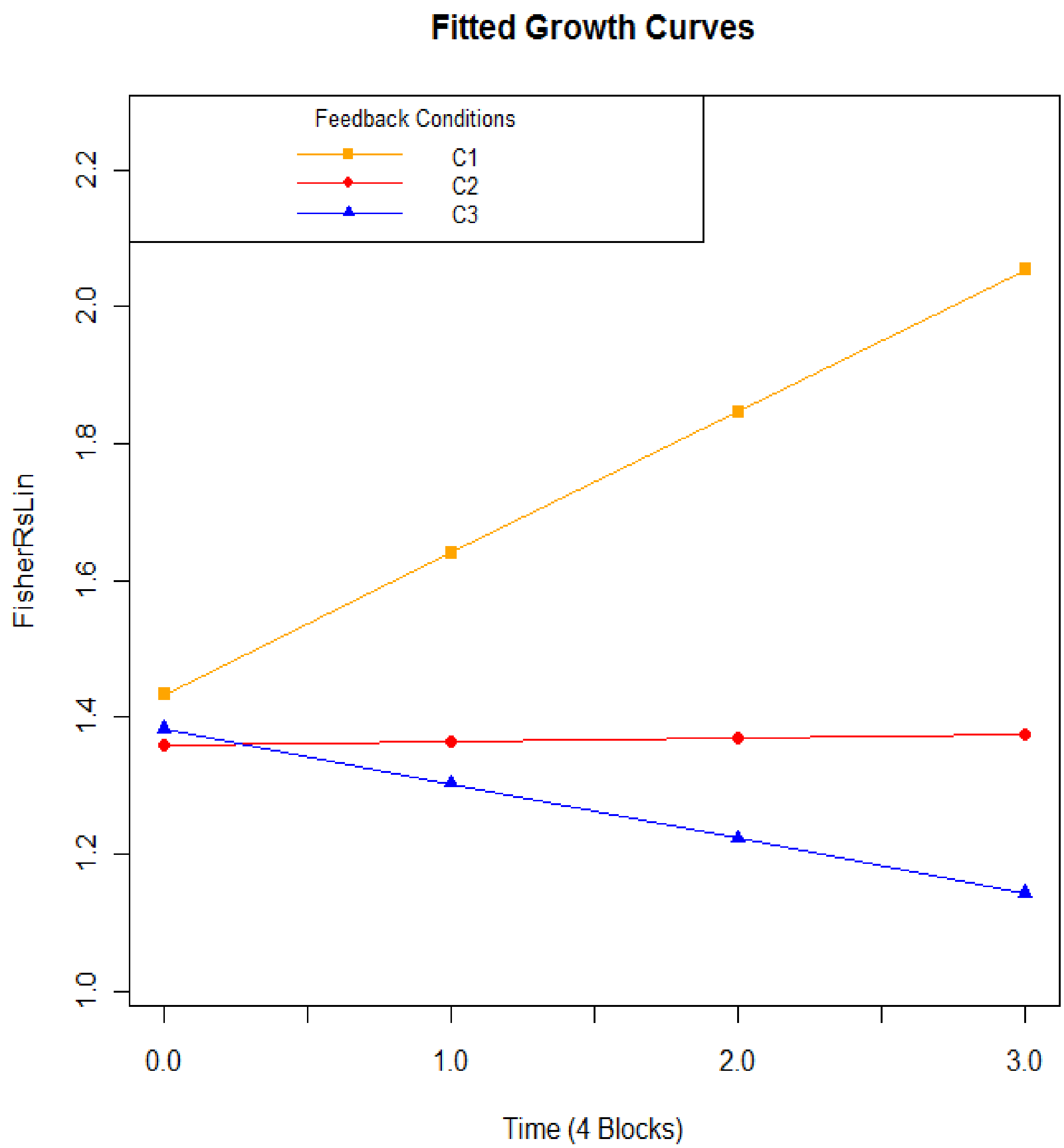


Figure 33. Change Over Time in C_{xj} (Relative Weight for the Disordinal Interaction): Trellis Plot of Slopes and Intercepts (N = All 142 Subjects) (Fall 2009)

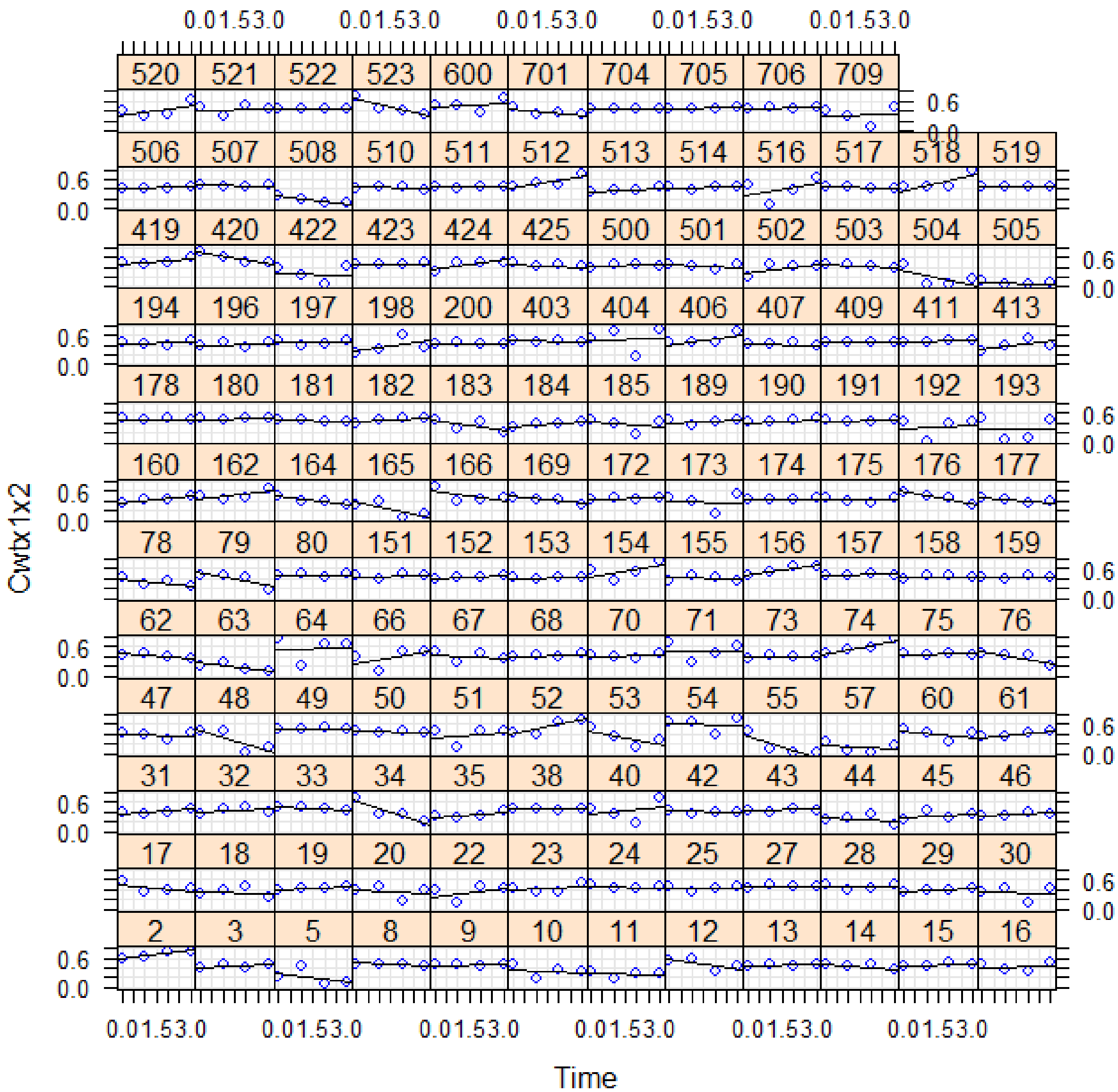


Figure 34. Change Over Time in C_{xj} (Relative Weight for the Disordinal Interaction): Boxplots of Slopes and Intercepts for Feedback Conditions (C1, C2, and C3) (Fall 2009)

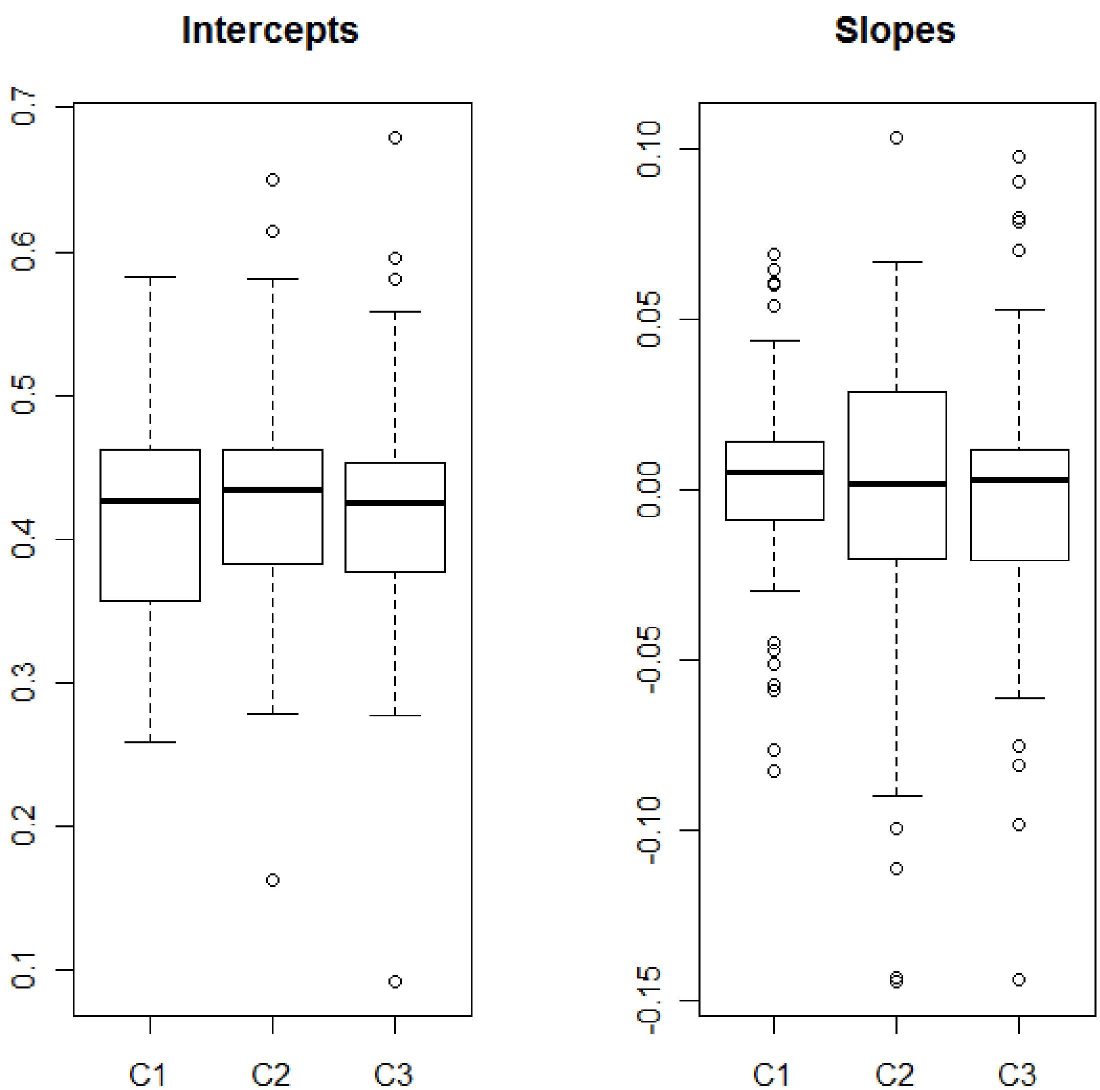


Figure 35. Change Over Time in C_{xj} (Relative Weight for the Disordinal Interaction): Subjects' Individual Regression Lines (N = All 142 Subjects) (Fall 2009)

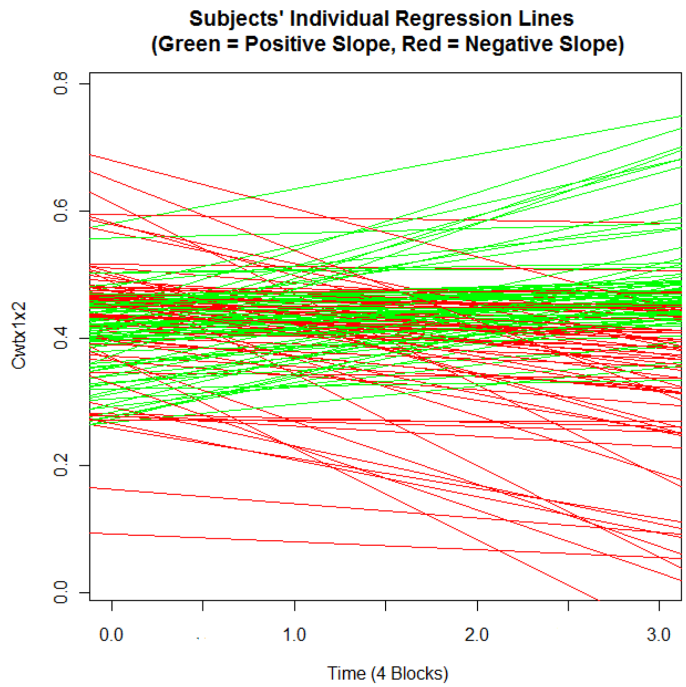


Figure 36. Change Over Time in C_{xj} (Relative Weight for the Disordinal Interaction): Fitted Growth Curve (Fixed Effect Only; Random Effects Ignored; No Interaction) (N = All 142 Subjects) (Fall 2009)

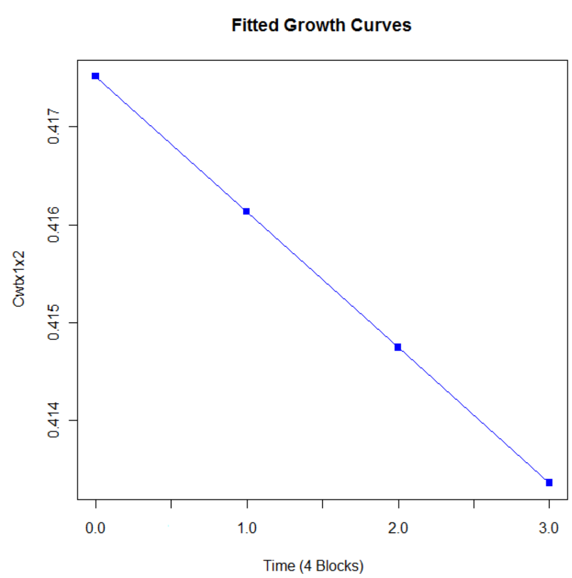


Figure 37. Change Over Time in Mean Absolute Confidence Levels for Each Feedback Condition (Based on Observed Sample-Level Data) (Fall 2009)

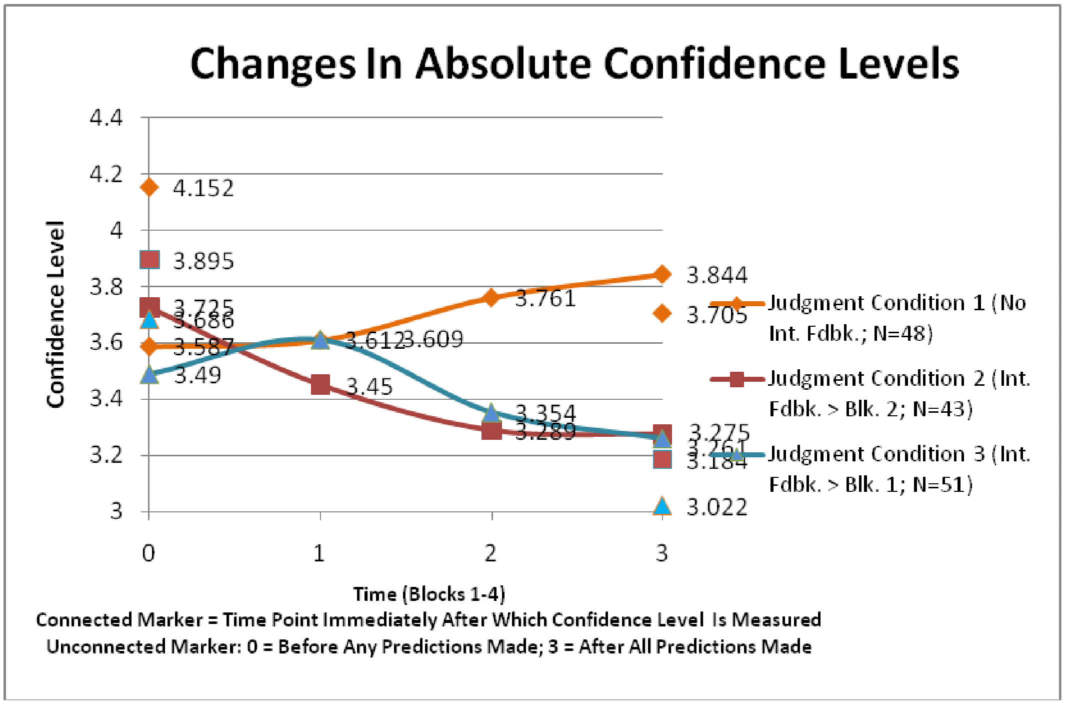


Figure 38. Change Over Time in Mean Relative Confidence Levels for Each Feedback Condition (Based on Observed Sample-Level Data) (Fall 2009)

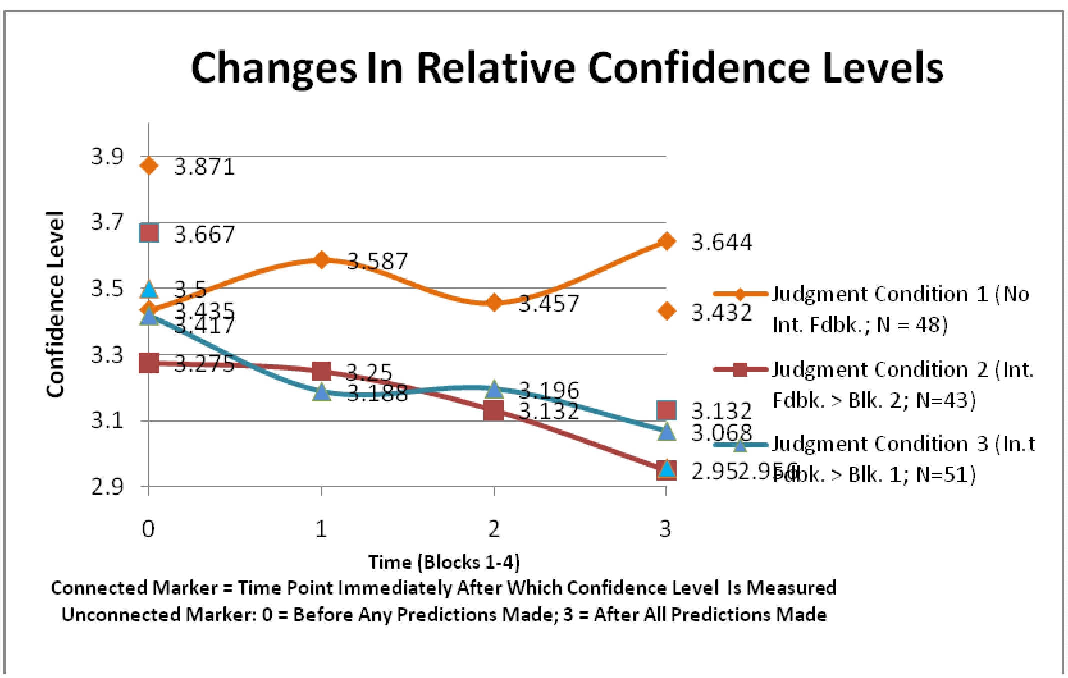


Figure 39. Change Over Time in Absolute Confidence, Trellis Plot of Slopes and Intercepts (N = 136 of 142 Subjects) (Fall 2009)

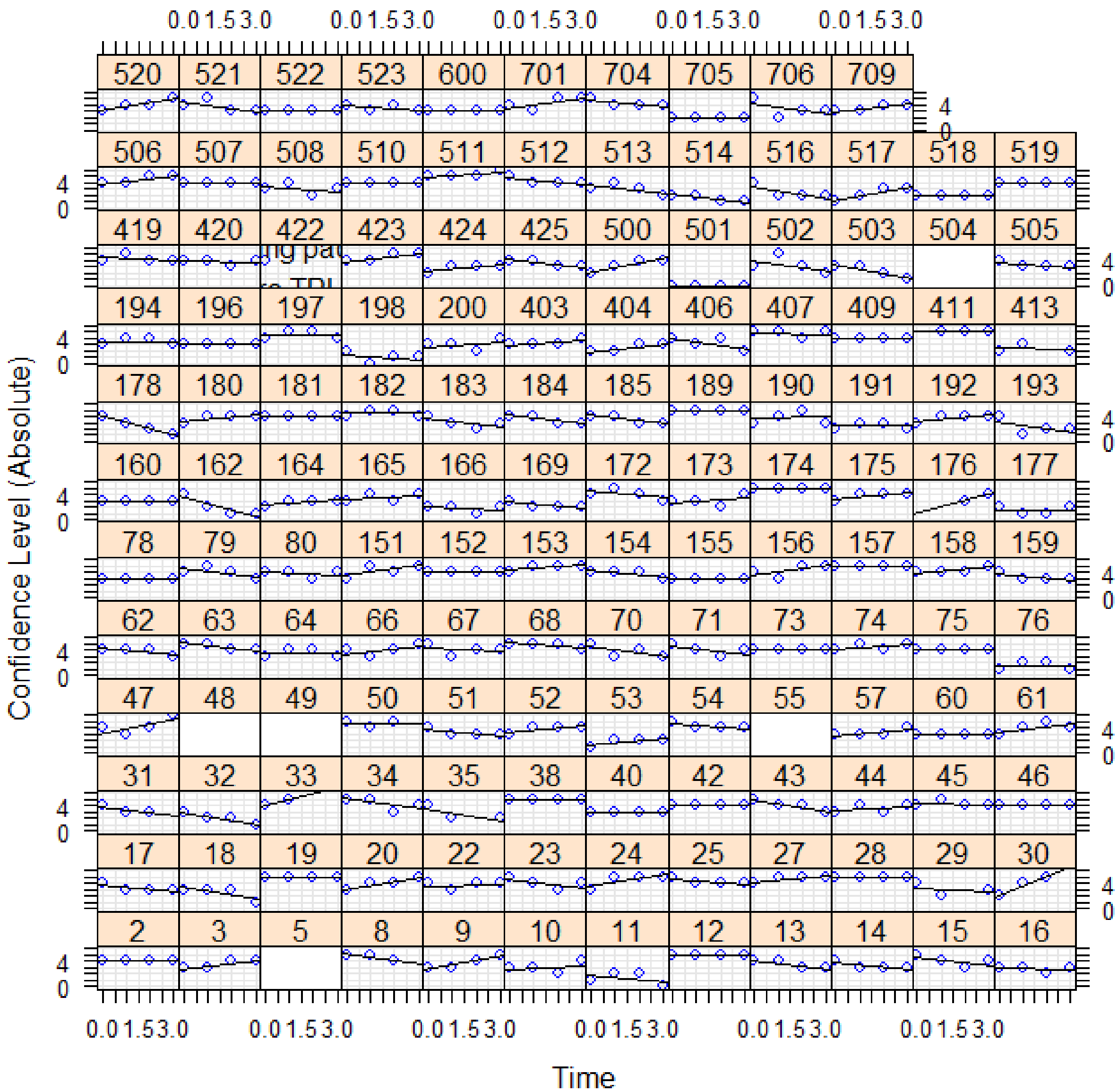


Figure 40. Change Over Time in Absolute Confidence, Boxplots of Slopes and Intercepts for Feedback Conditions (C1, C2, and C3) (Fall 2009)

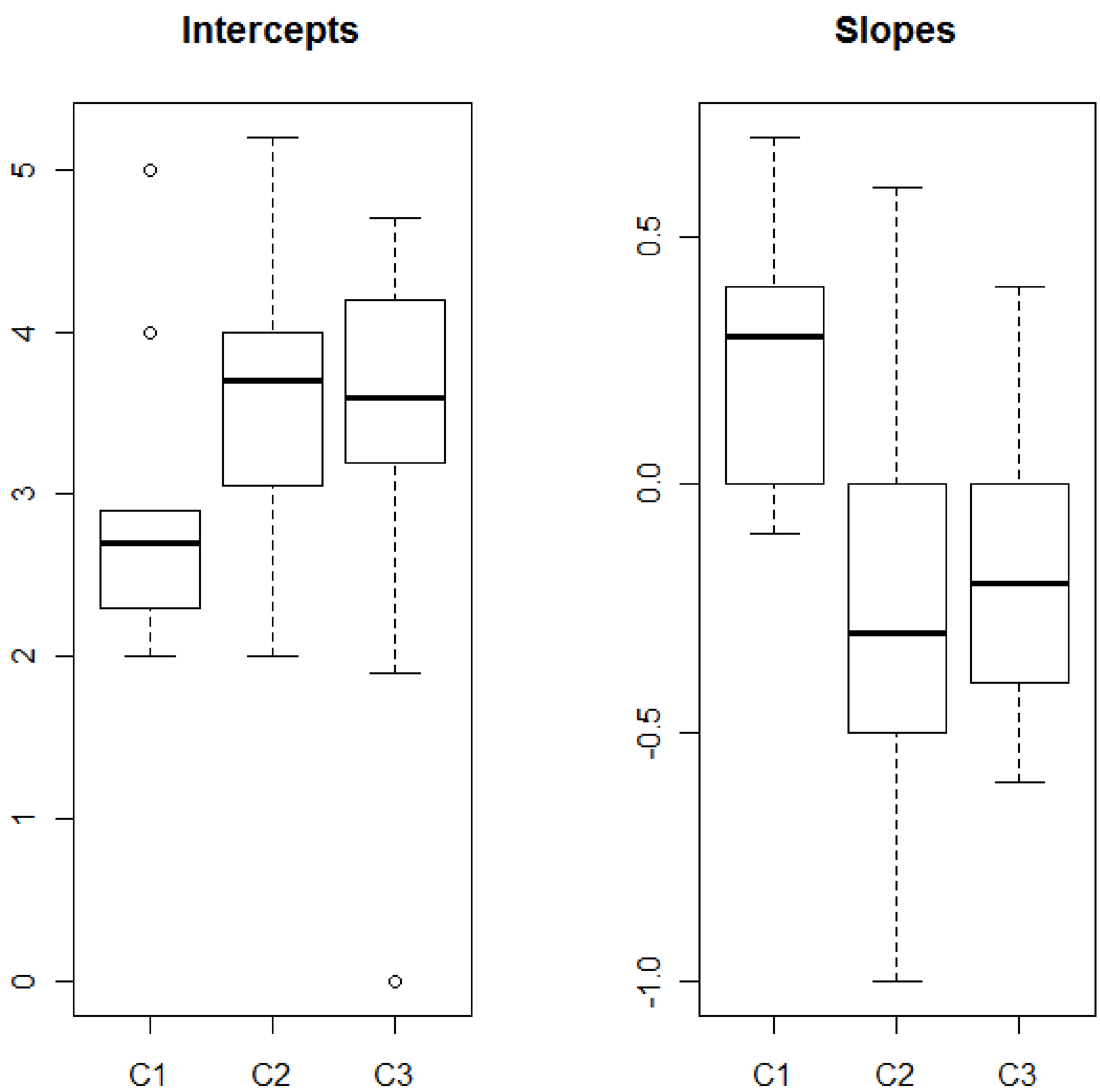


Figure 41. Change Over Time in Absolute Confidence: Fitted Growth Curves (Fall 2009)

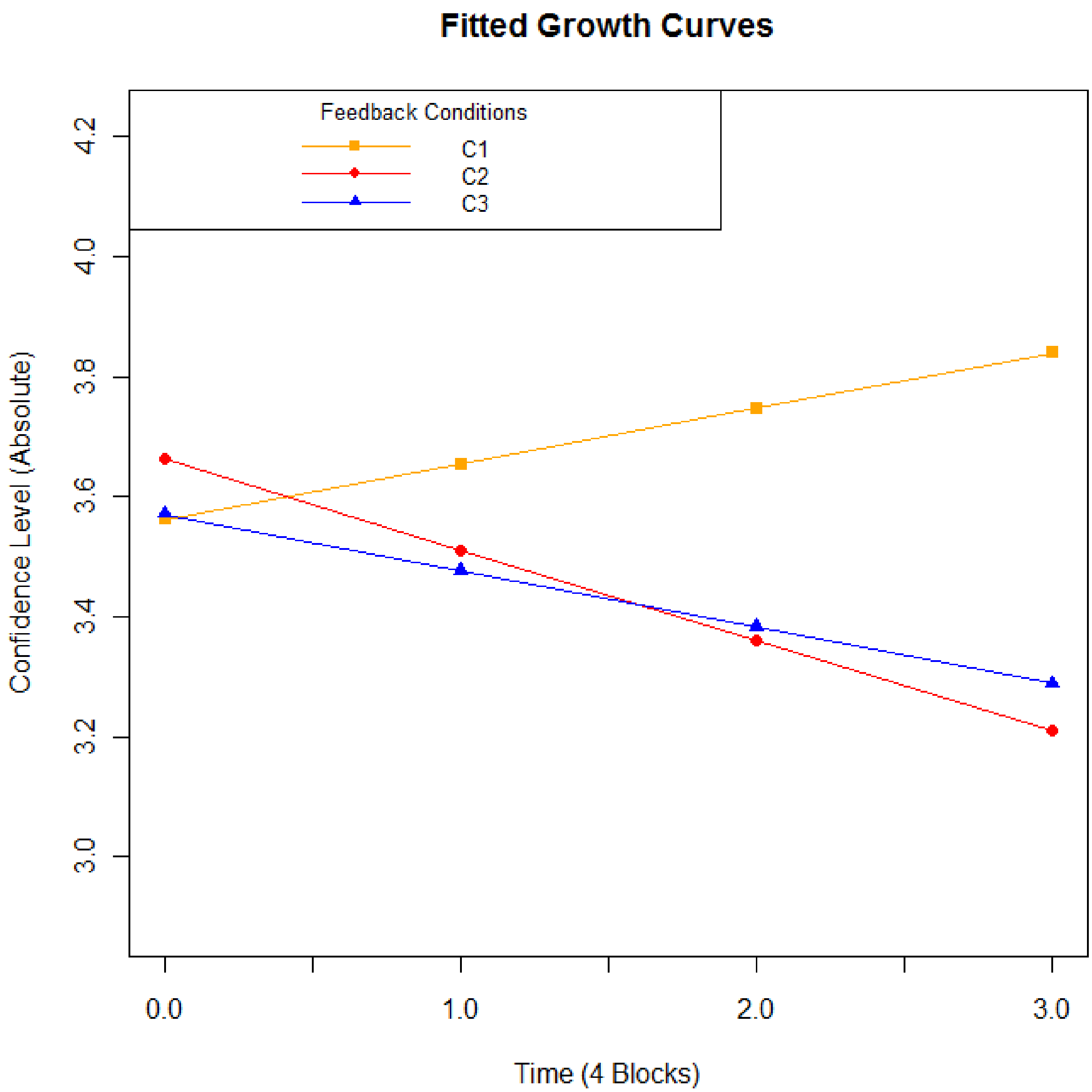


Figure 42. Change Over Time in Relative Confidence: Trellis Plot of Slopes and Intercepts (*N* = 135 of 142 Subjects) (Fall 2009)

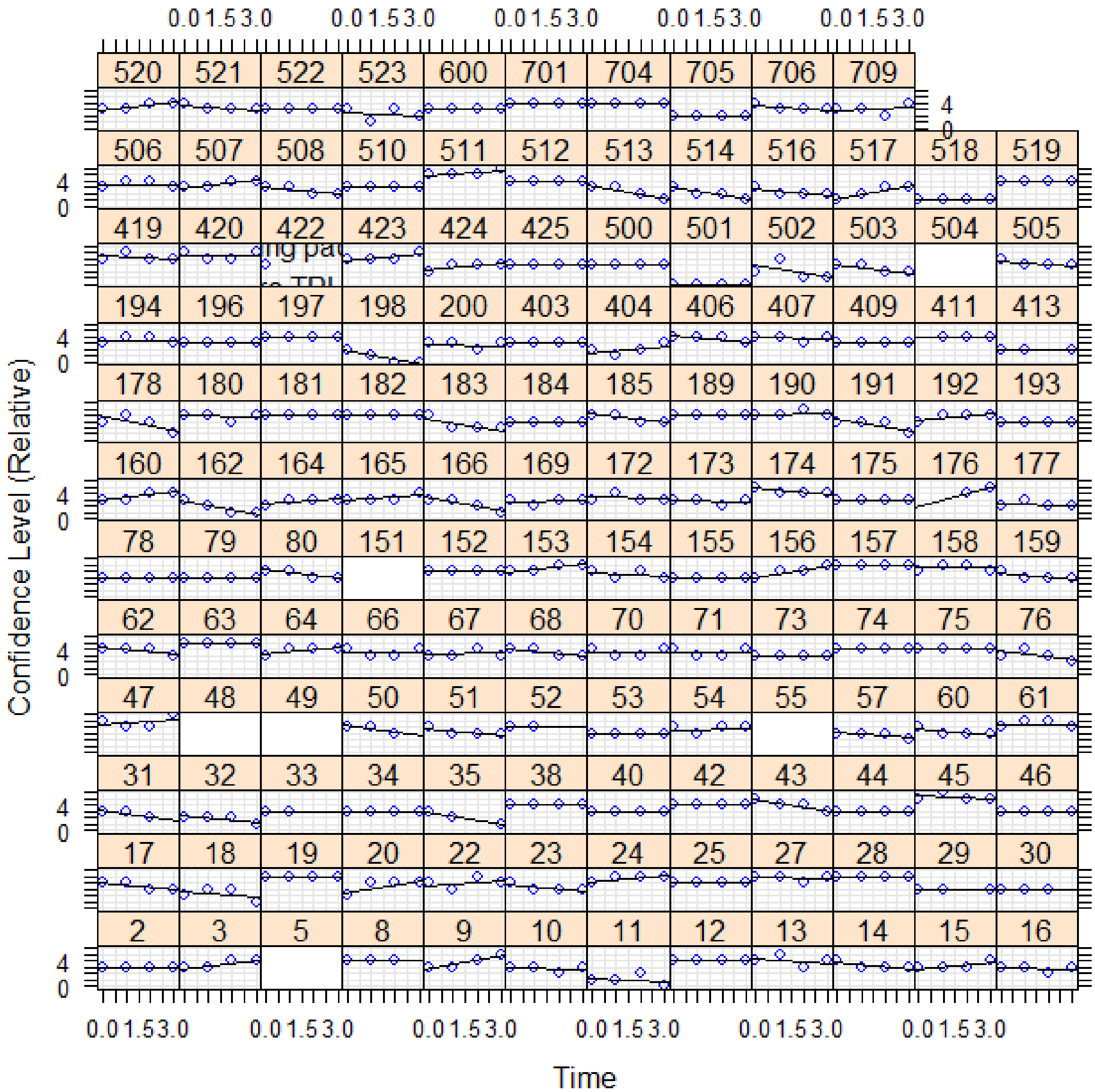


Figure 43. Change Over Time in Relative Confidence: Boxplots of Slopes and Intercepts for Feedback Conditions (C1, C2, and C3) (Fall 2009)

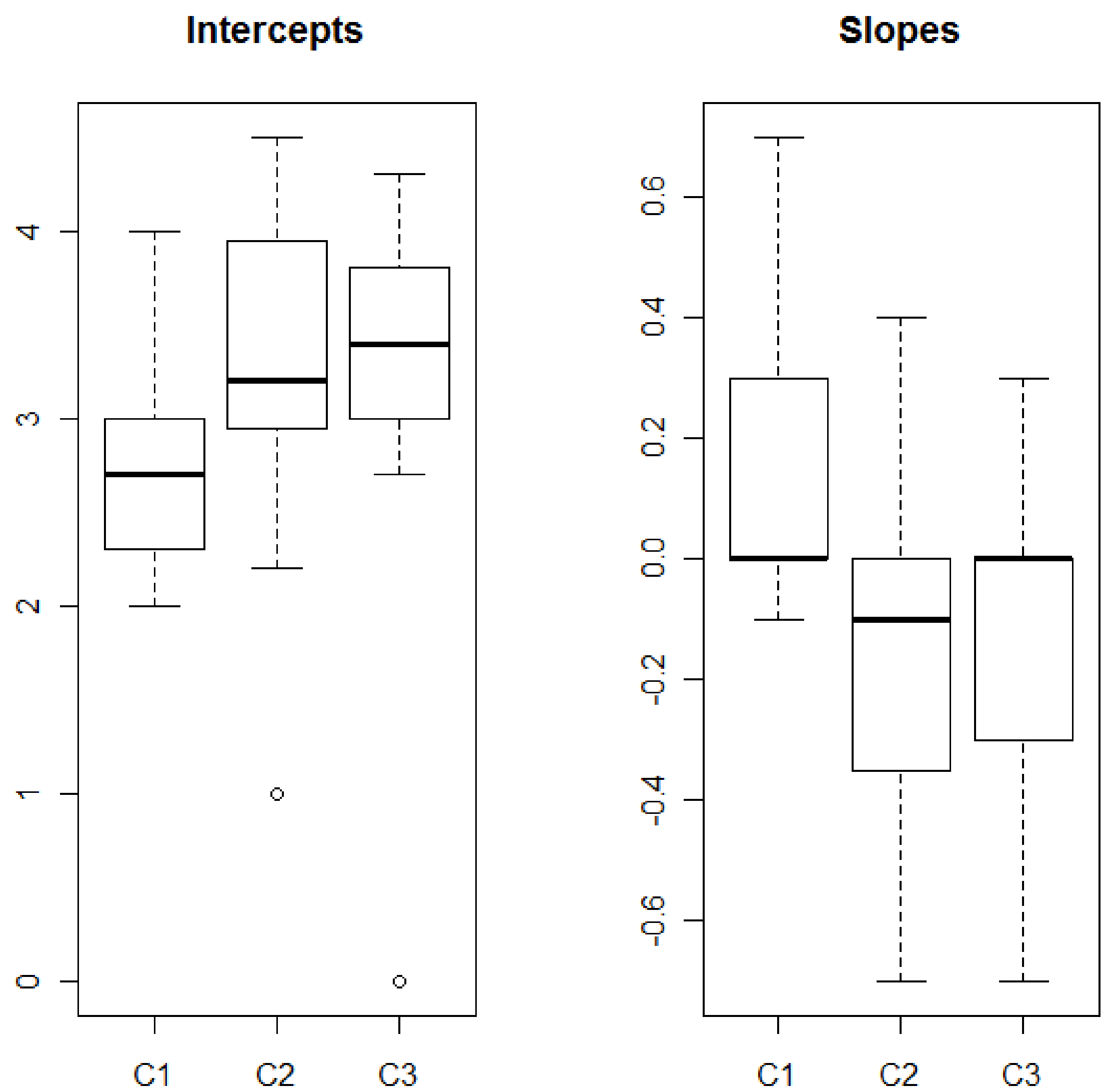


Figure 44. Change Over Time in Relative Confidence: Fitted Fixed Effect Growth Curve (No Random Effects; Without Interaction) (N = 135 of 142 Subjects) (Fall 2009)

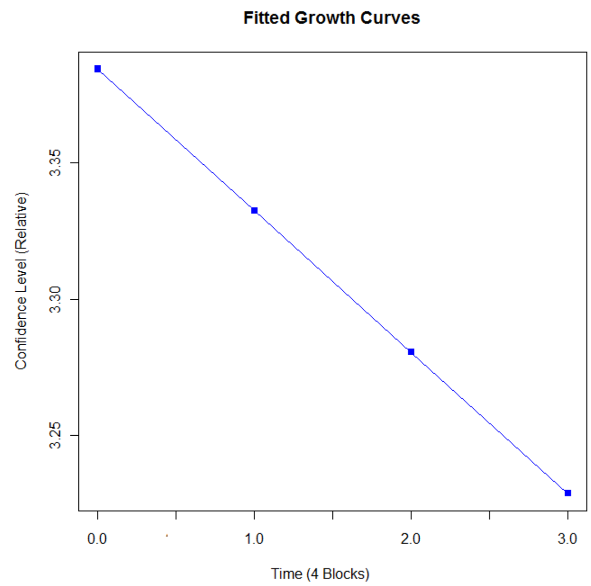


Figure 45. Change Over Time in Relative Confidence: Fitted Growth Curves for Each Feedback Condition (Fall 2009)

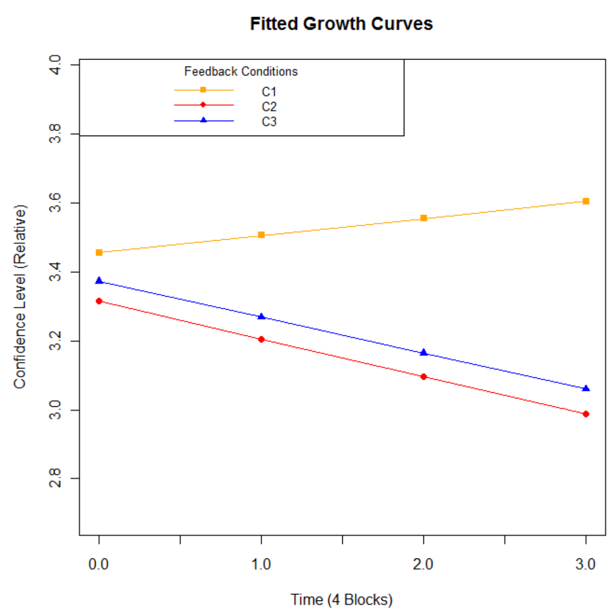


Figure 46. Prediction Accuracy After Training: r_a (Spring 2010)

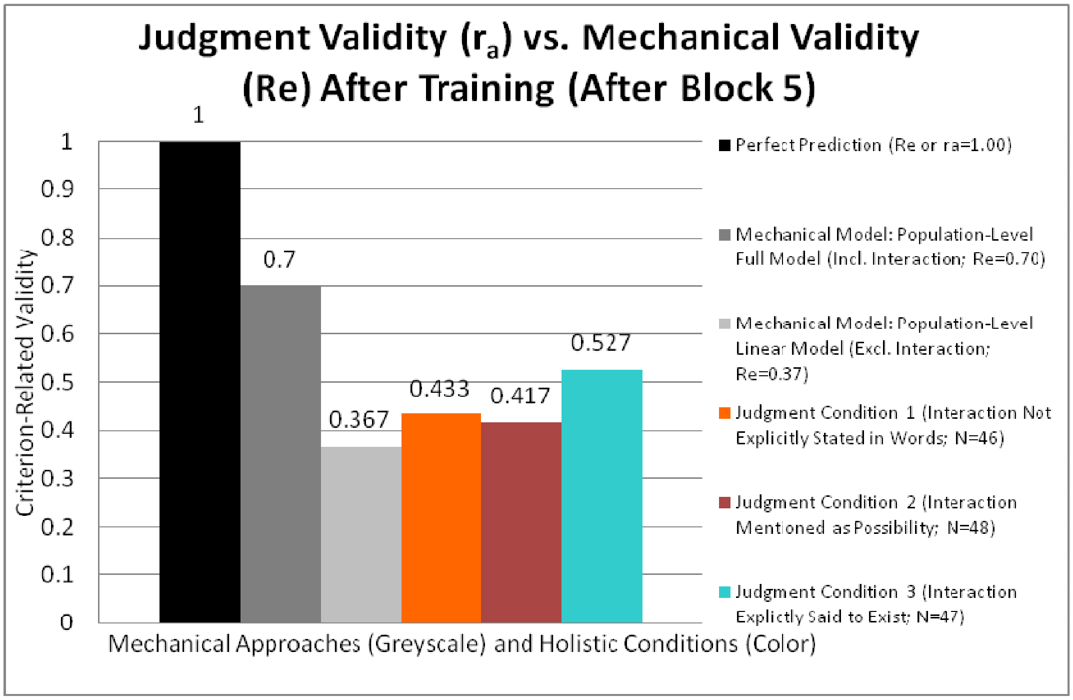


Figure 47. Prediction Accuracy After Training: Skill Score (Spring 2010)

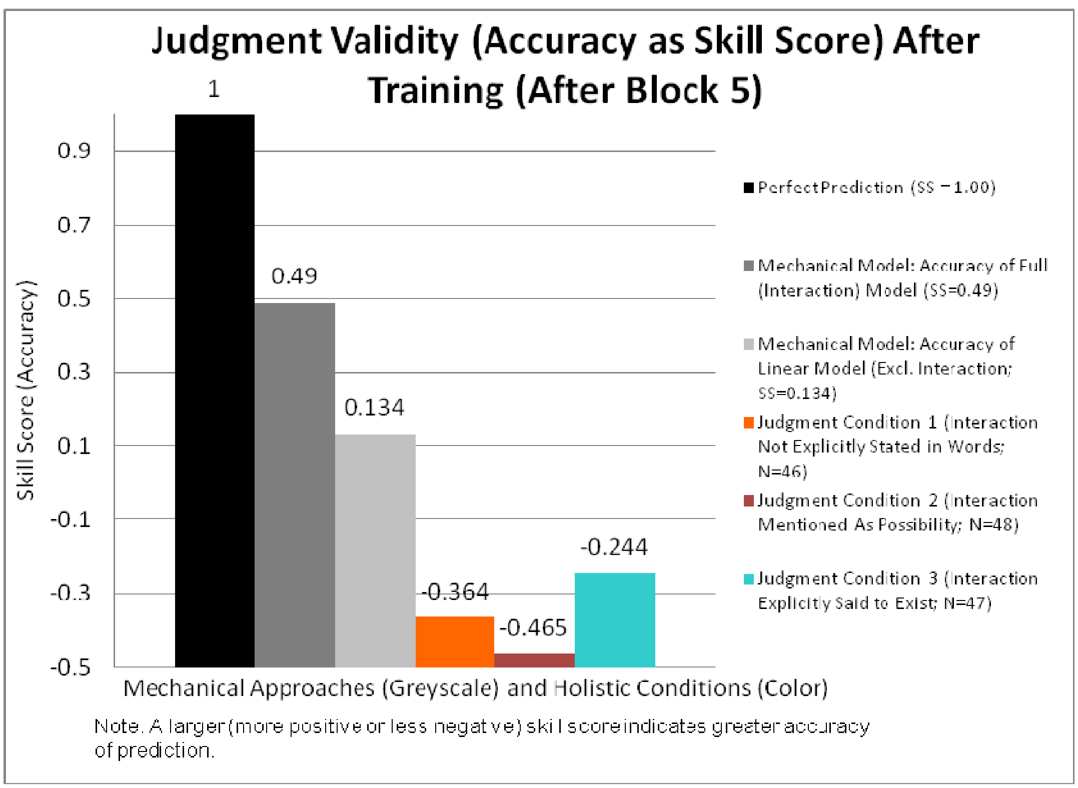


Figure 48. Prediction Accuracy During Training: r_a (Spring 2010)

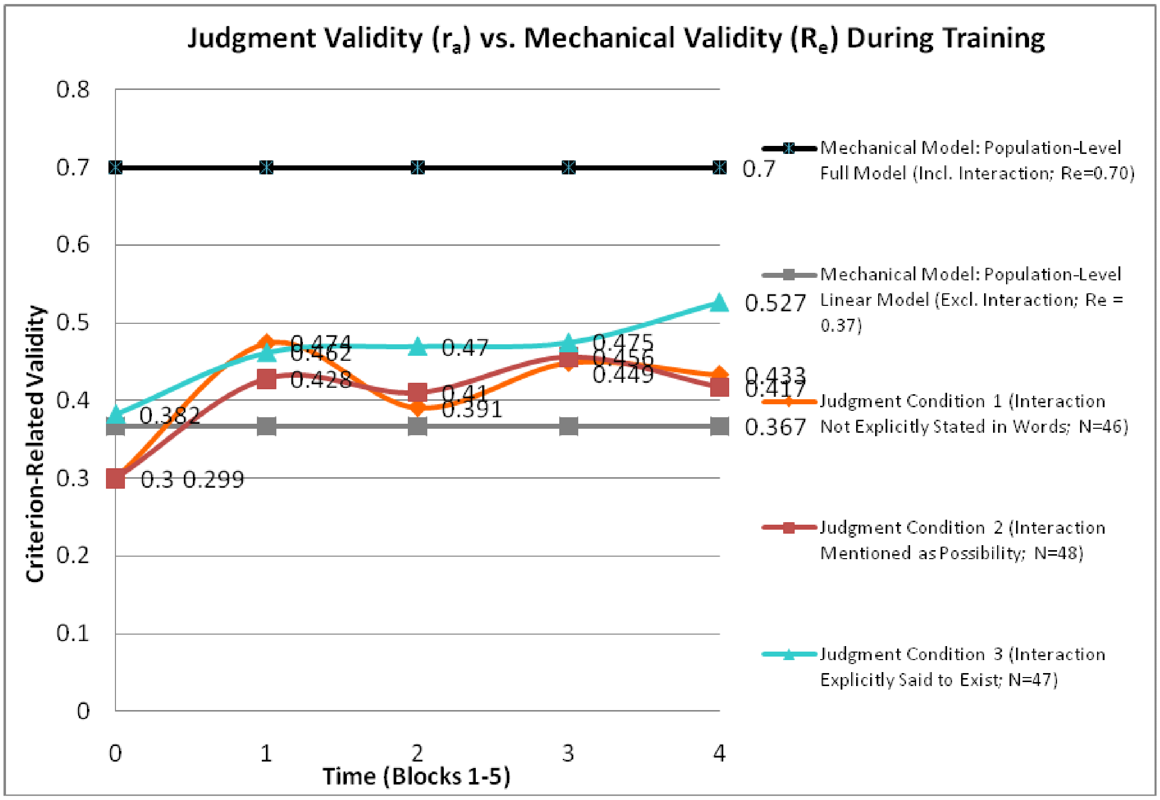


Figure 49. Prediction Accuracy During Training: Skill Score (Spring 2010)

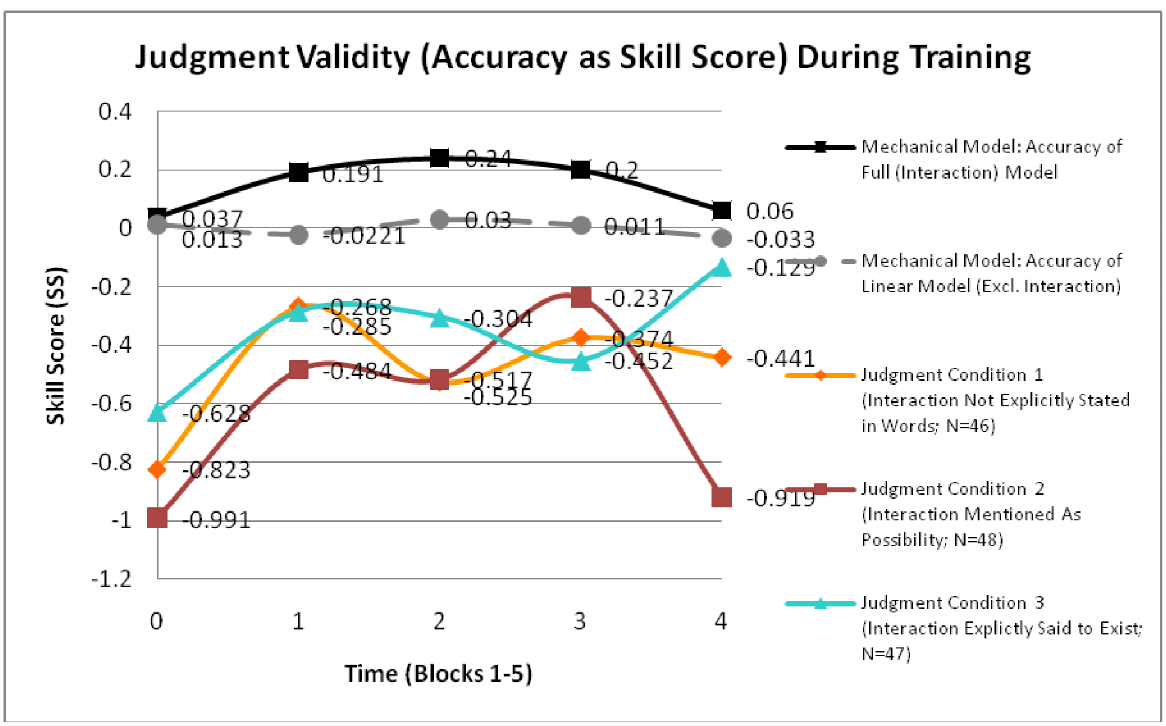


Figure 50. Change Over Time in Fisher r_a (Judgment Criterion-Related Validity): Boxplots of Slopes and Intercepts for Feedback Conditions (C1, C2, and C3) (Spring 2010)

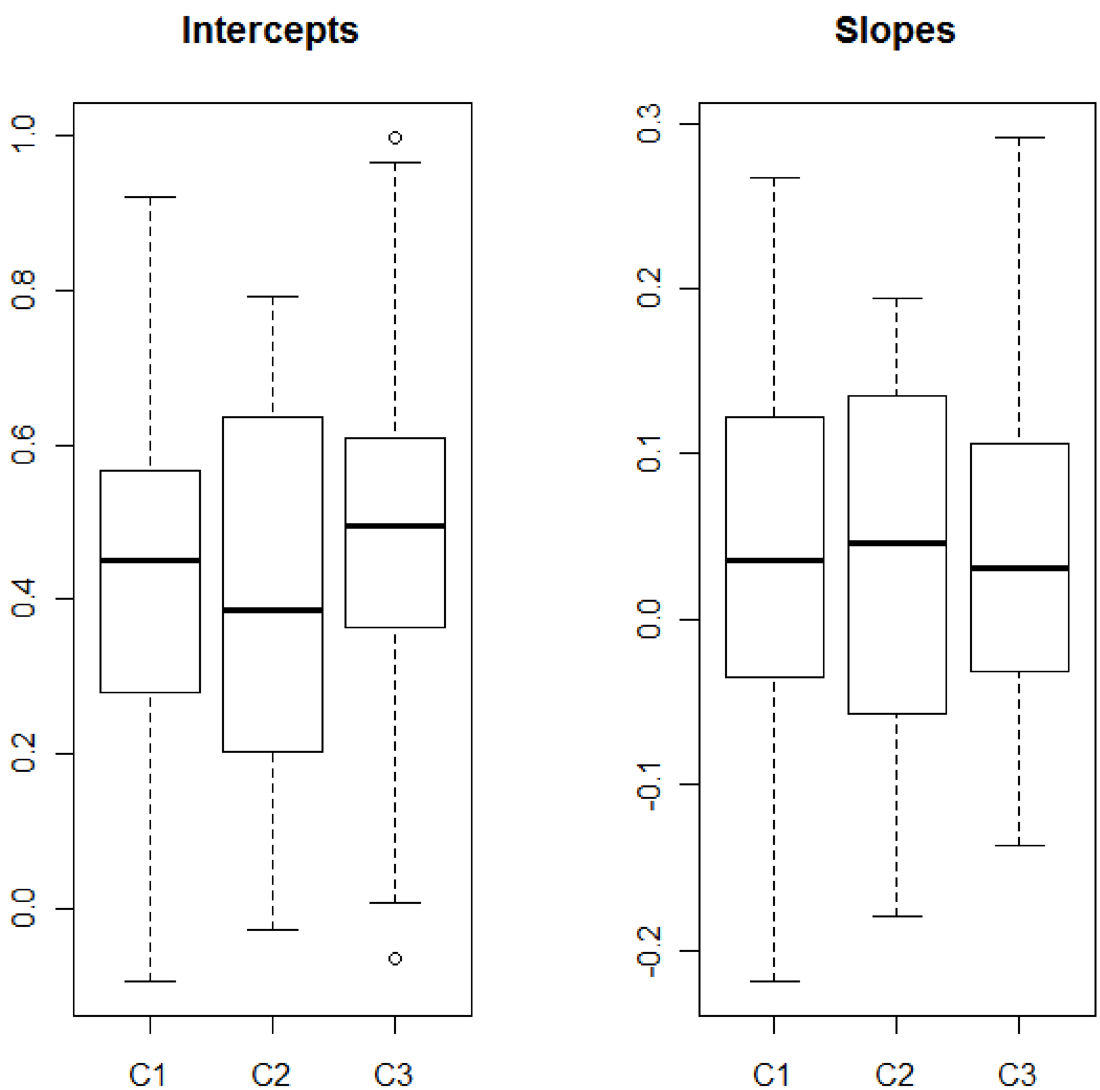


Figure 51. Change Over Time in Fisher r_a (Judgment Criterion-Related Validity): Fitted Growth Curve If Feedback x Time Interaction is Ignored (Spring 2010)

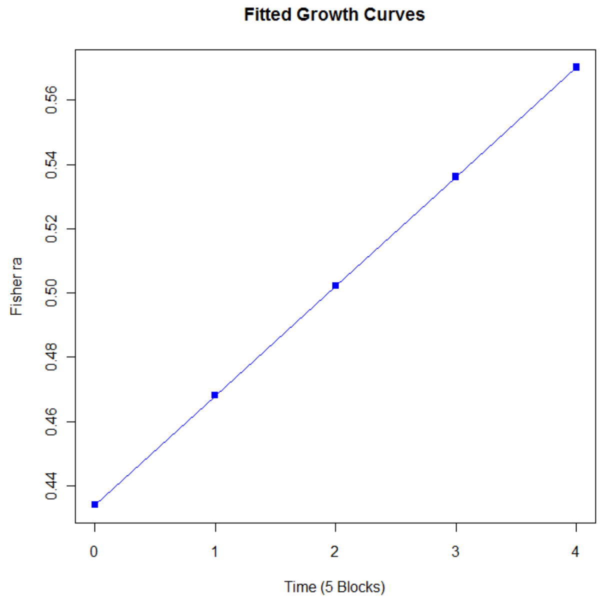


Figure 52. Change Over Time in Fisher r_a (Judgment Criterion-Related Validity): Fitted Growth Curves if Feedback x Time Interaction is Considered (Spring 2010)

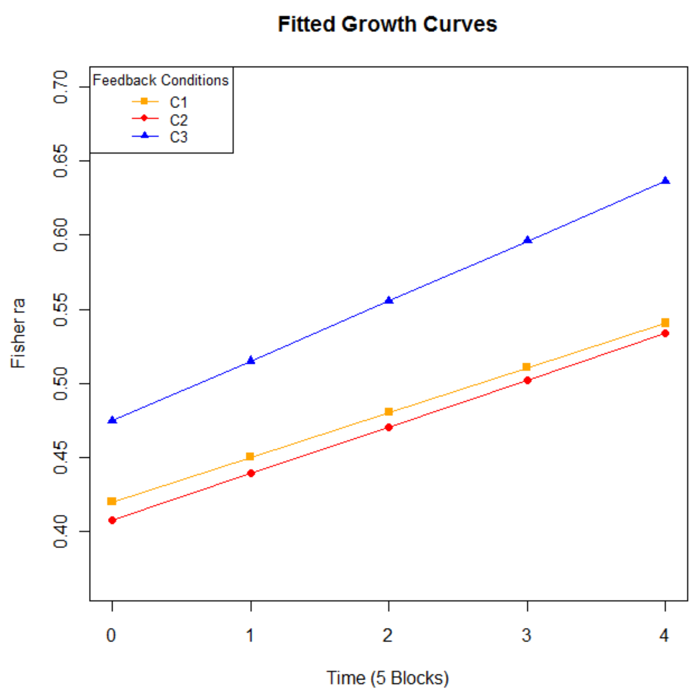


Figure 53. Change Over Time in Skill Score (Accuracy): Boxplots of Slopes and Intercepts for Feedback Conditions (C1, C2, and C3) (Spring 2010)

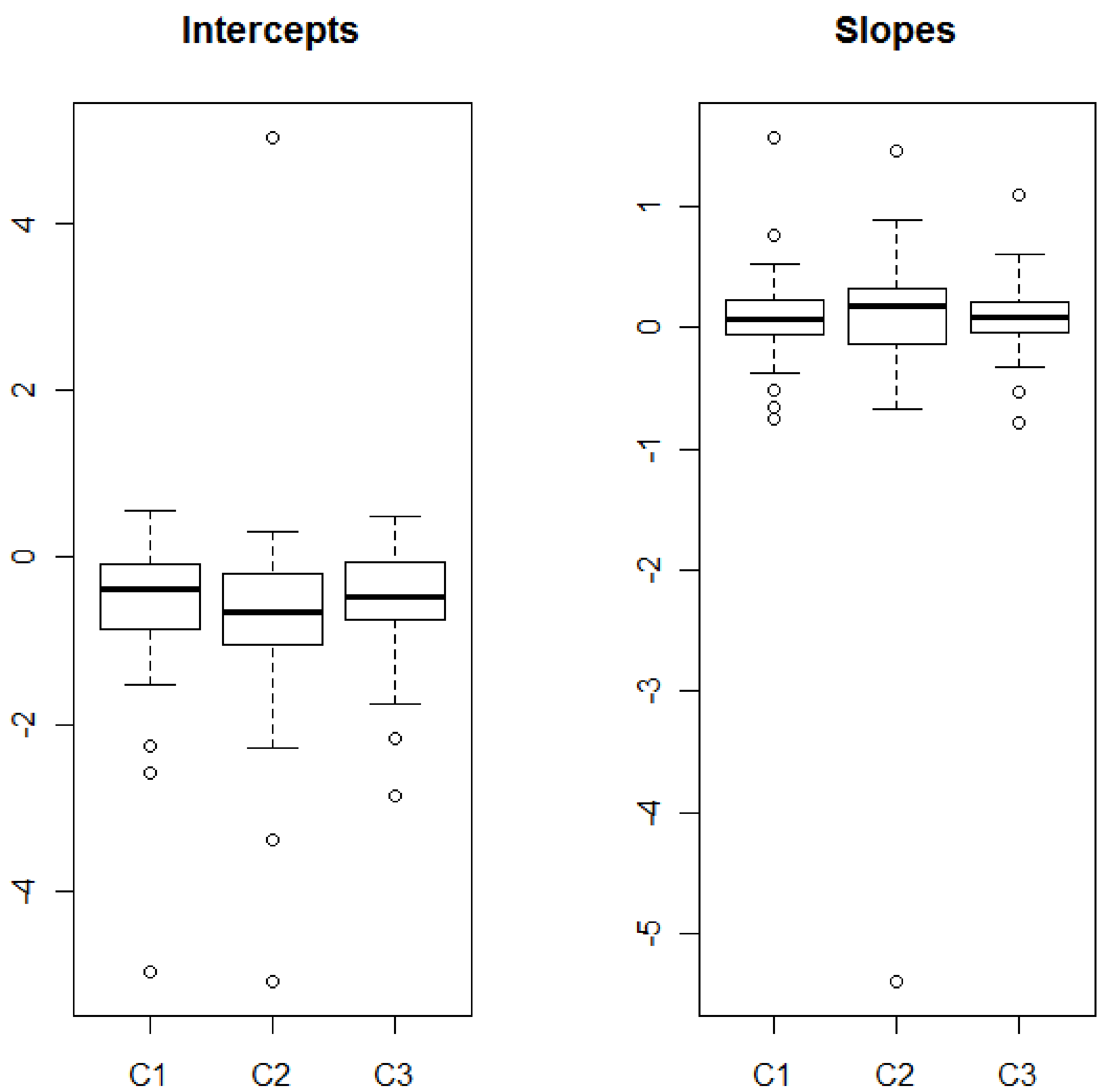


Figure 54. Change Over Time in Skill Score (Accuracy): Fitted Growth Curve (Spring 2010)

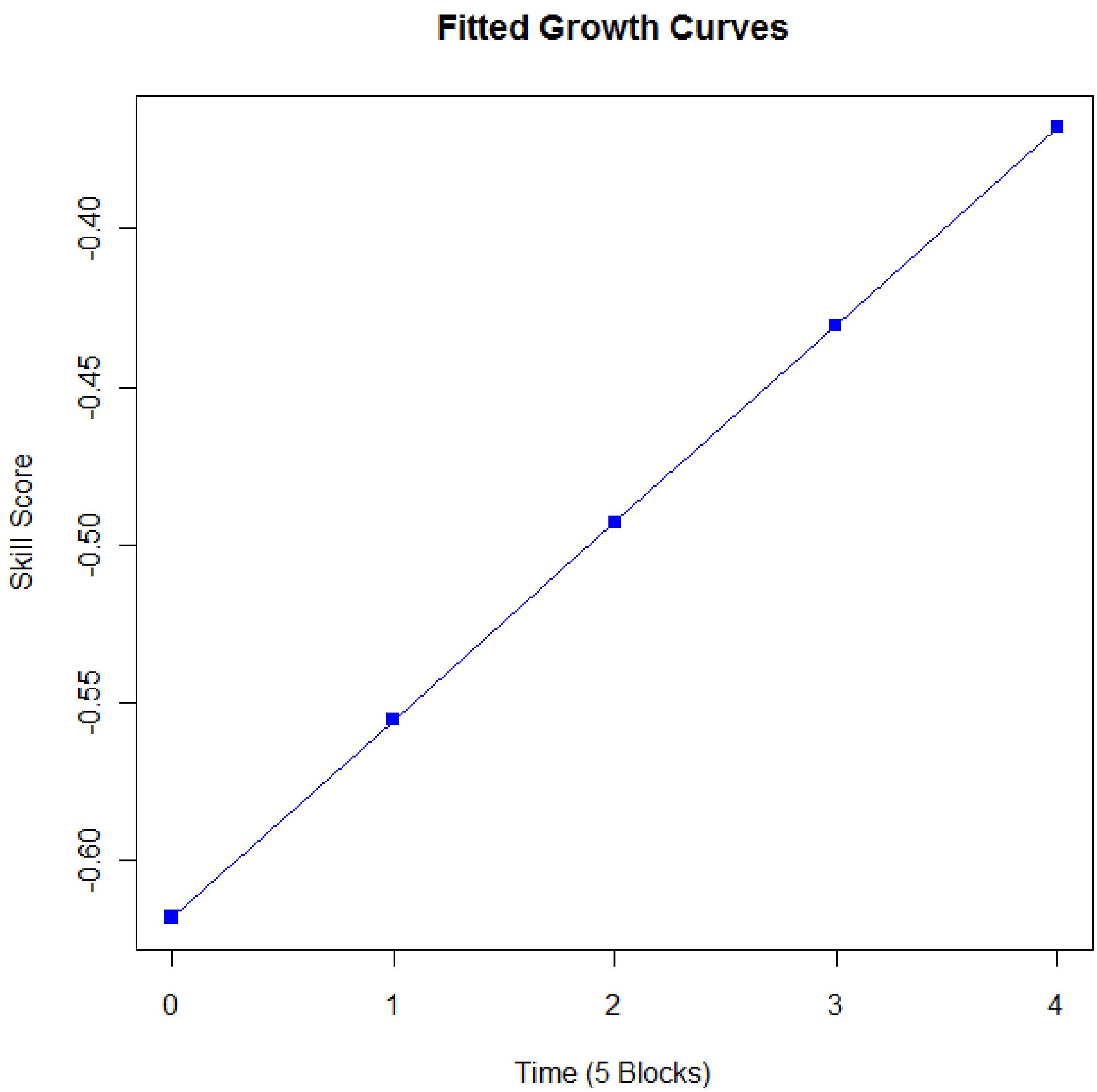


Figure 55. Change Over Time in Fisher *C* (Unmodeled Knowledge): Trellis Plot of Slopes and Intercepts (*N* = All 141 Subjects) (Spring 2010)

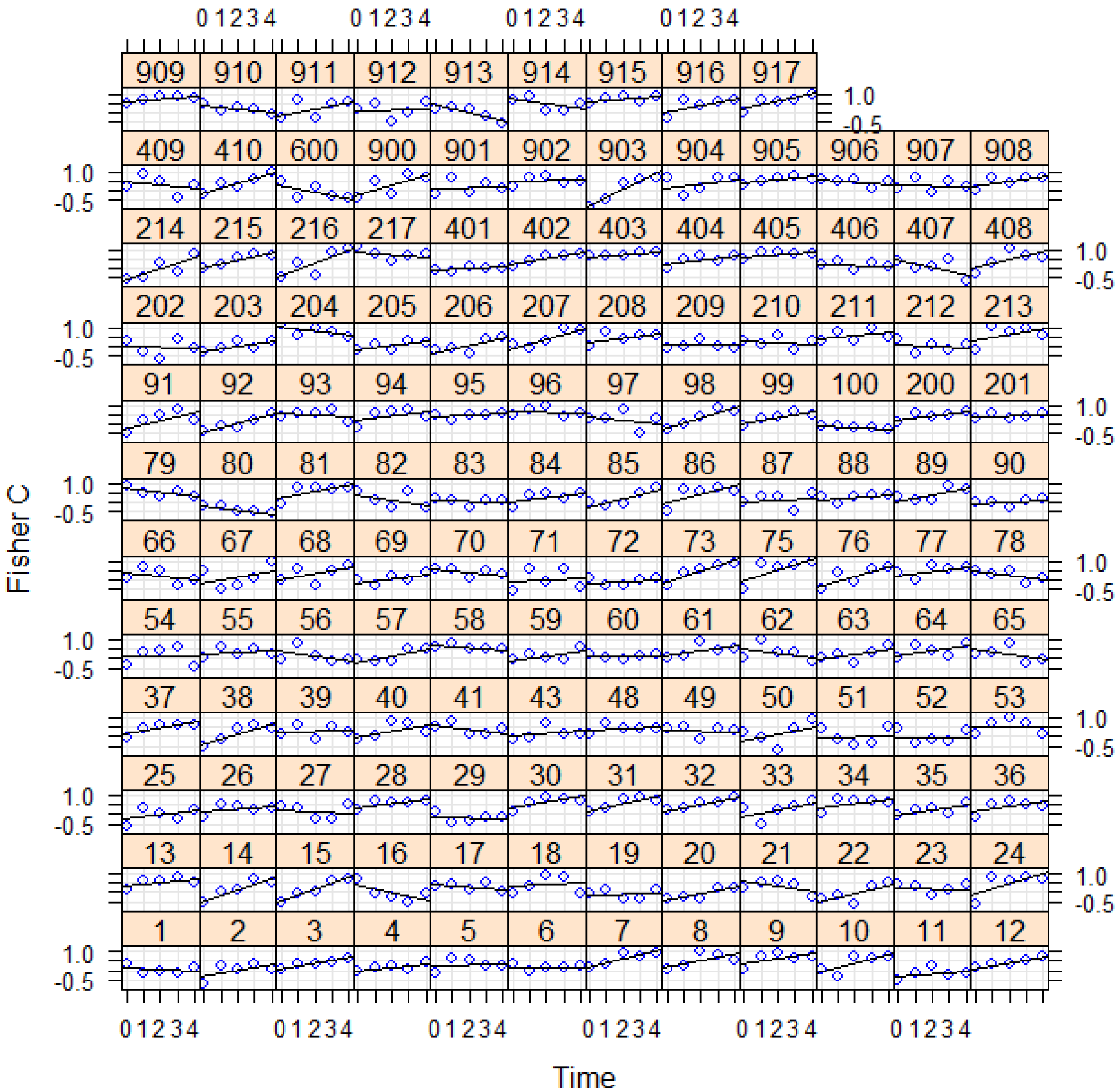


Figure 56. Change Over Time in Fisher *C* (Unmodeled Knowledge): Boxplots of Slopes and Intercepts for Feedback Conditions (C1, C2, and C3) (Spring 2010)

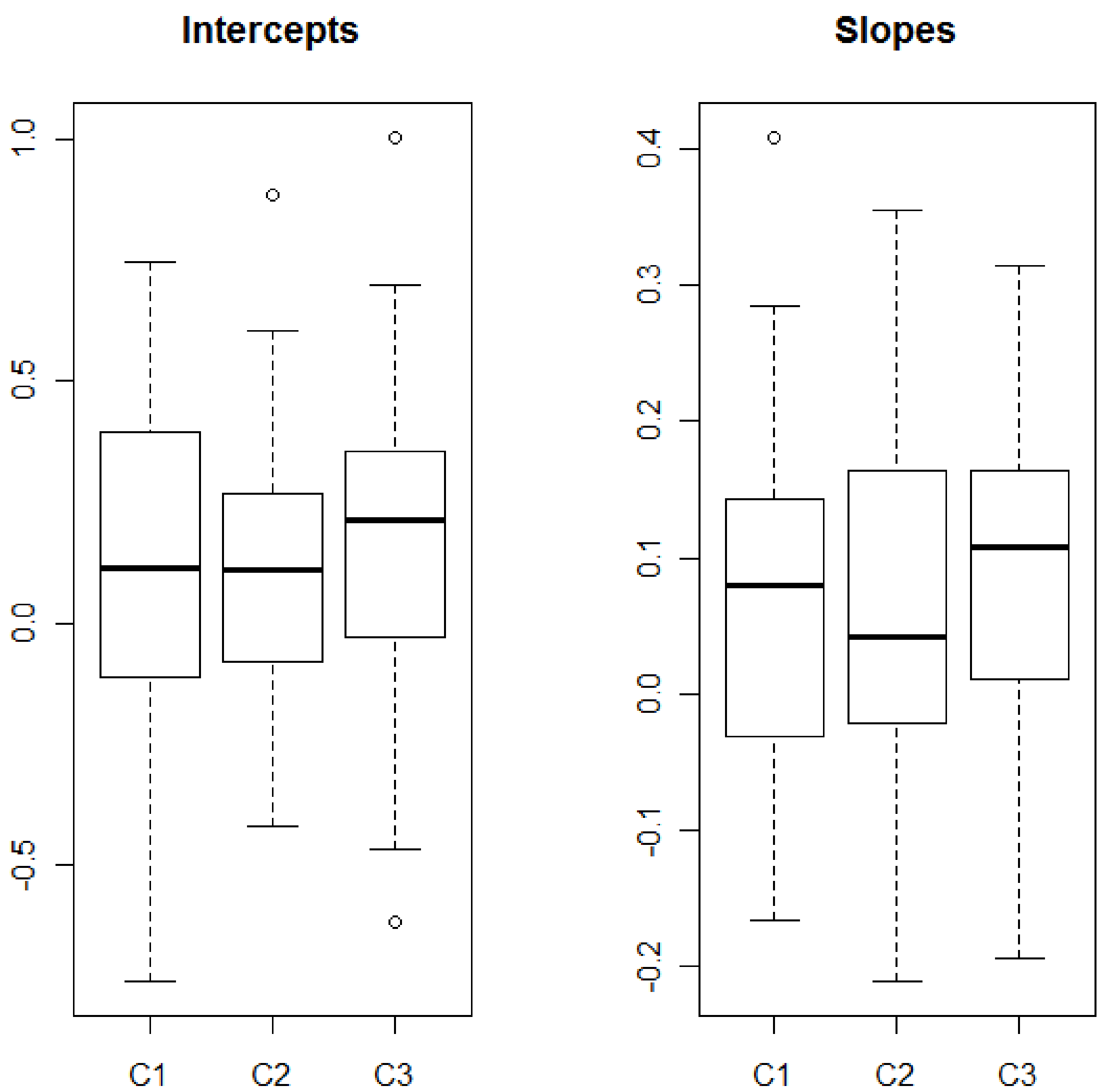


Figure 57. Change Over Time in Fisher *C* (Unmodeled Knowledge): Subjects' Individual Regression Lines (*N* = All 141 Subjects) (Spring 2010)

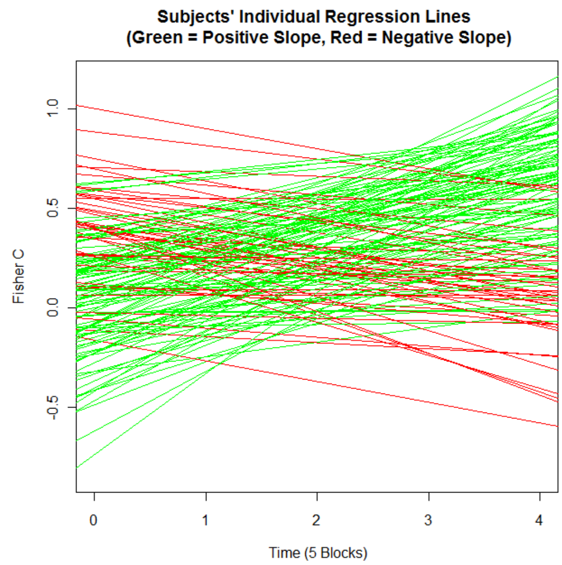


Figure 58. Change Over Time in Fisher *C* (Unmodeled Knowledge): Regression Line For Fixed Effect (Random Effects Ignored; With No Interaction)

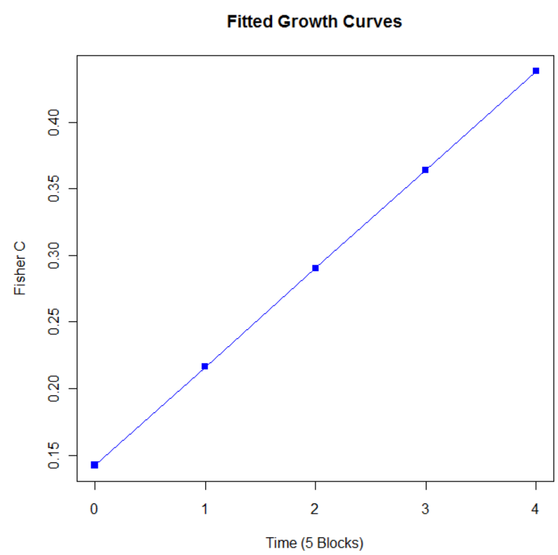


Figure 59. Change Over Time in Fisher r_z (Criterion-Related Validity of Unmodeled Knowledge): Trellis Plot of Slopes and Intercepts (N = All 141 Subjects) (Spring 2010)

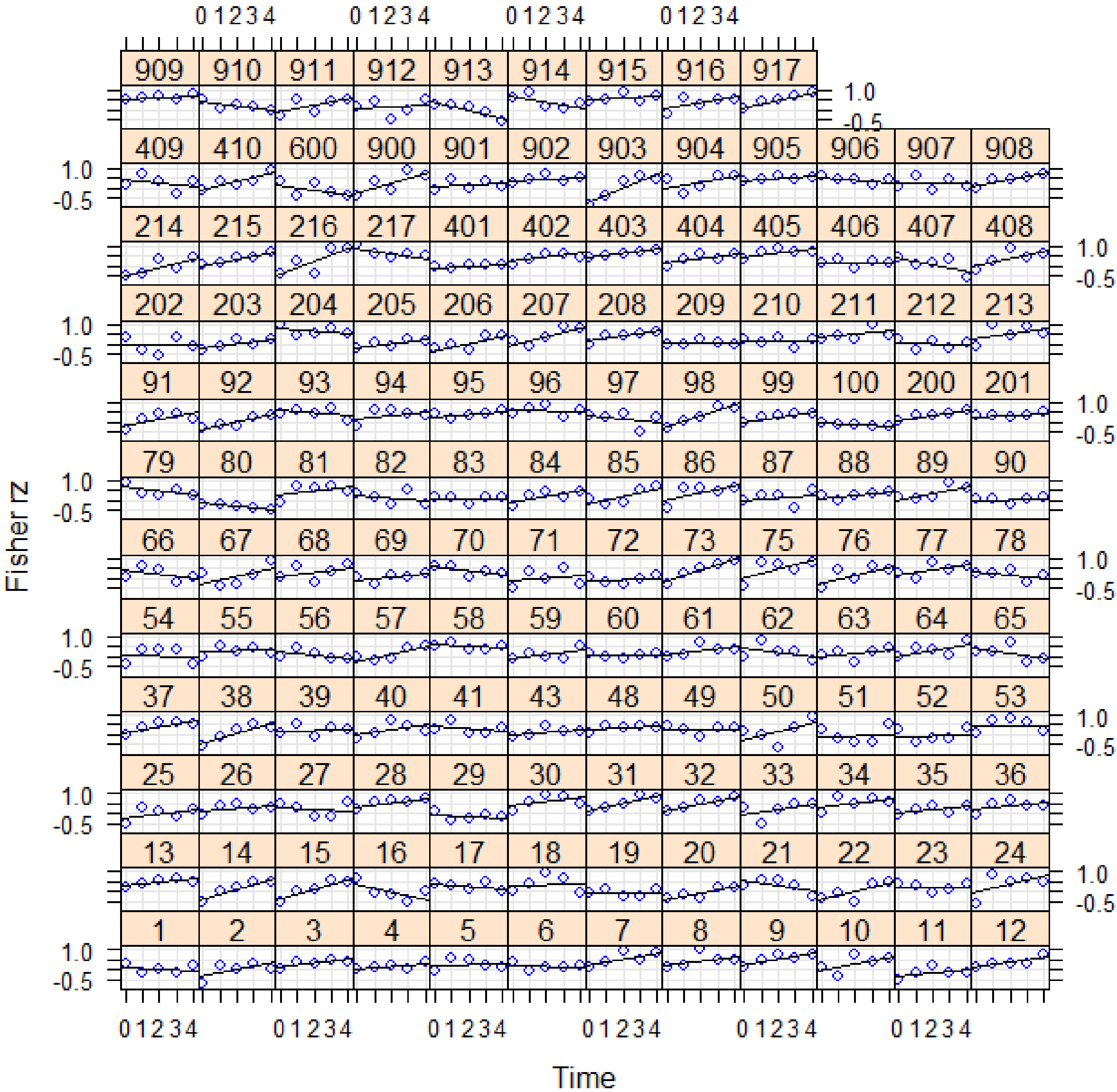


Figure 60. Change Over Time in Fisher r_z (Criterion-Related Validity of Unmodeled Knowledge): Boxplots of Slopes and Intercepts for Feedback Conditions (C1, C2, and C3) (Spring 2010)

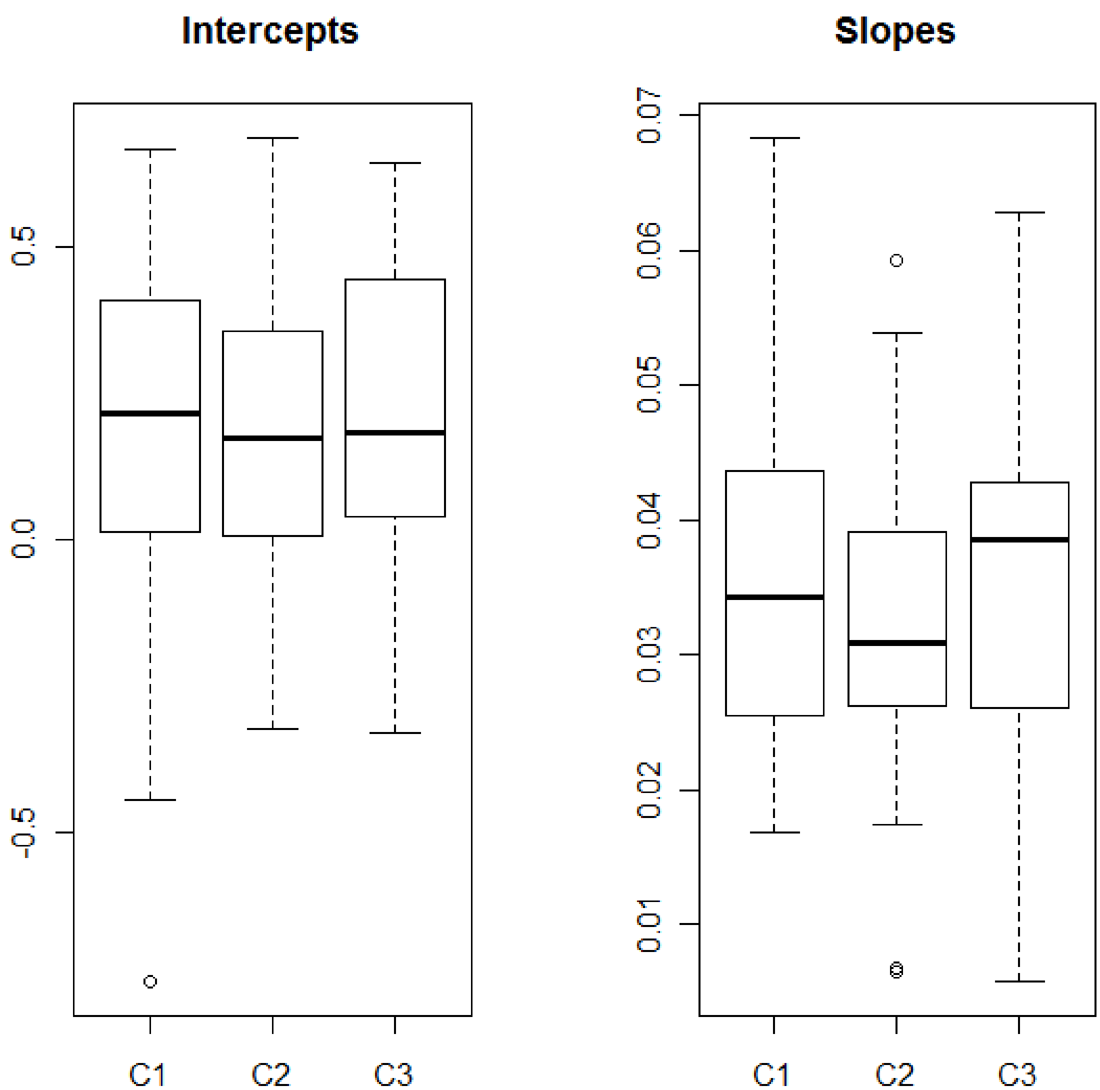


Figure 61. Change Over Time in Fisher r_z (Criterion-Related Validity of Unmodeled Knowledge): Fitted Fixed Effect Growth Curve (No Random Effects; Without Interaction) (*N* = All 141 Subjects) (Spring 2010)

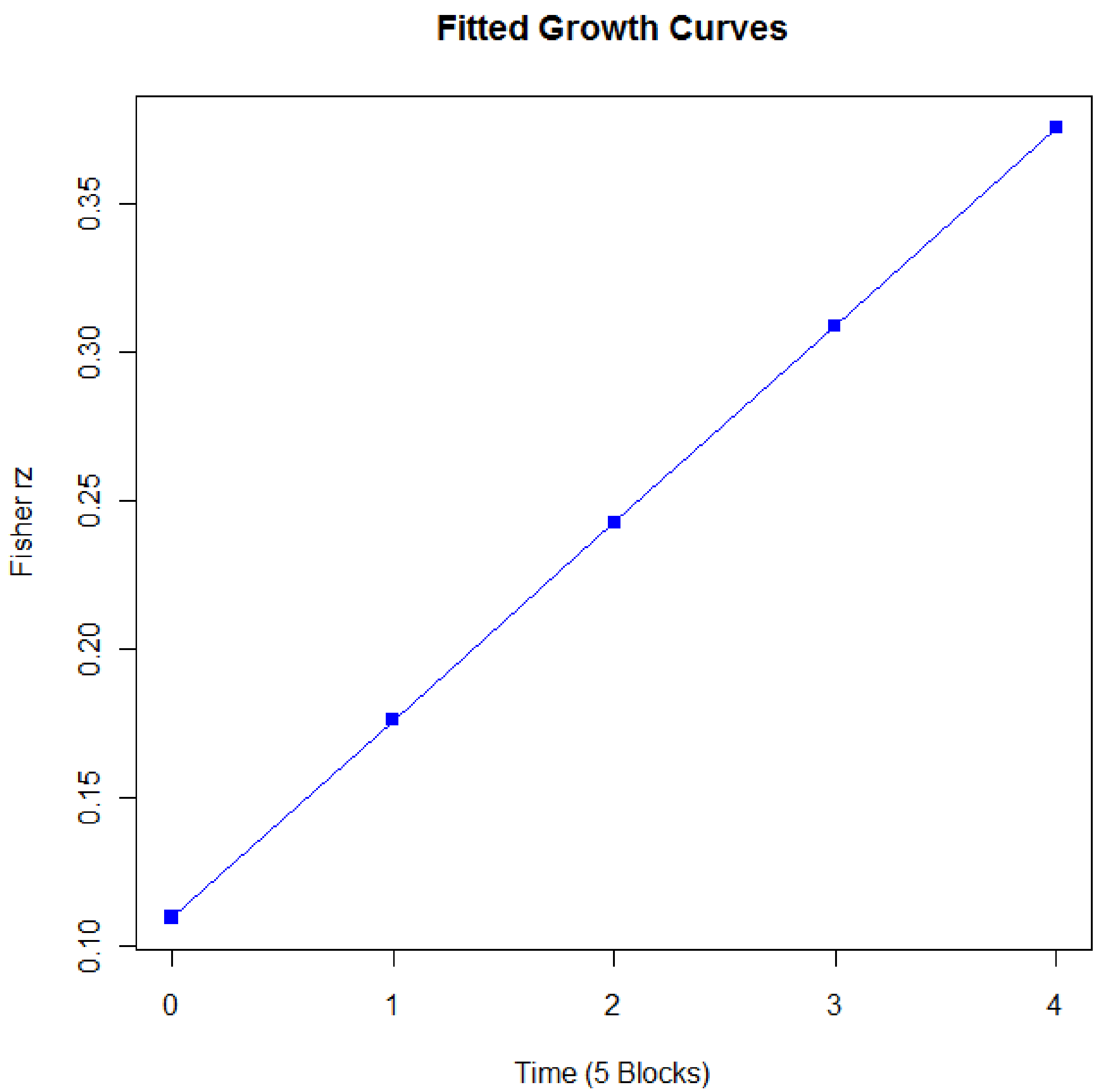


Figure 62. Change Over Time in Fisher *G* (Mechanical Knowledge): Boxplots of Slopes and Intercepts for Feedback Conditions (C1, C2, and C3) (Spring 2010)

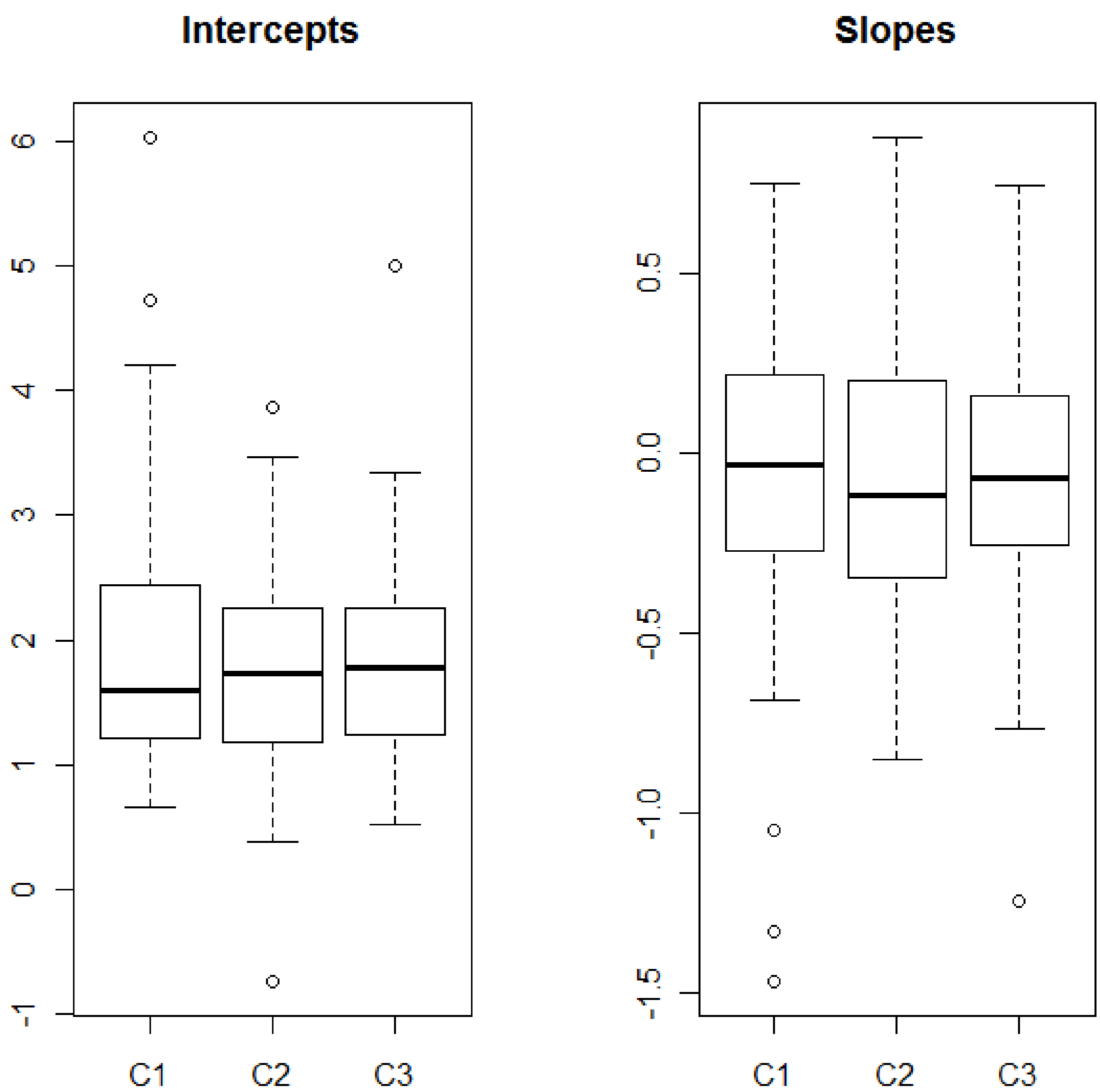


Figure 63. Change Over Time in Fisher *G* (Mechanical Knowledge): Fitted Growth Curve (*N* = All 141 Subjects) (Spring 2010)

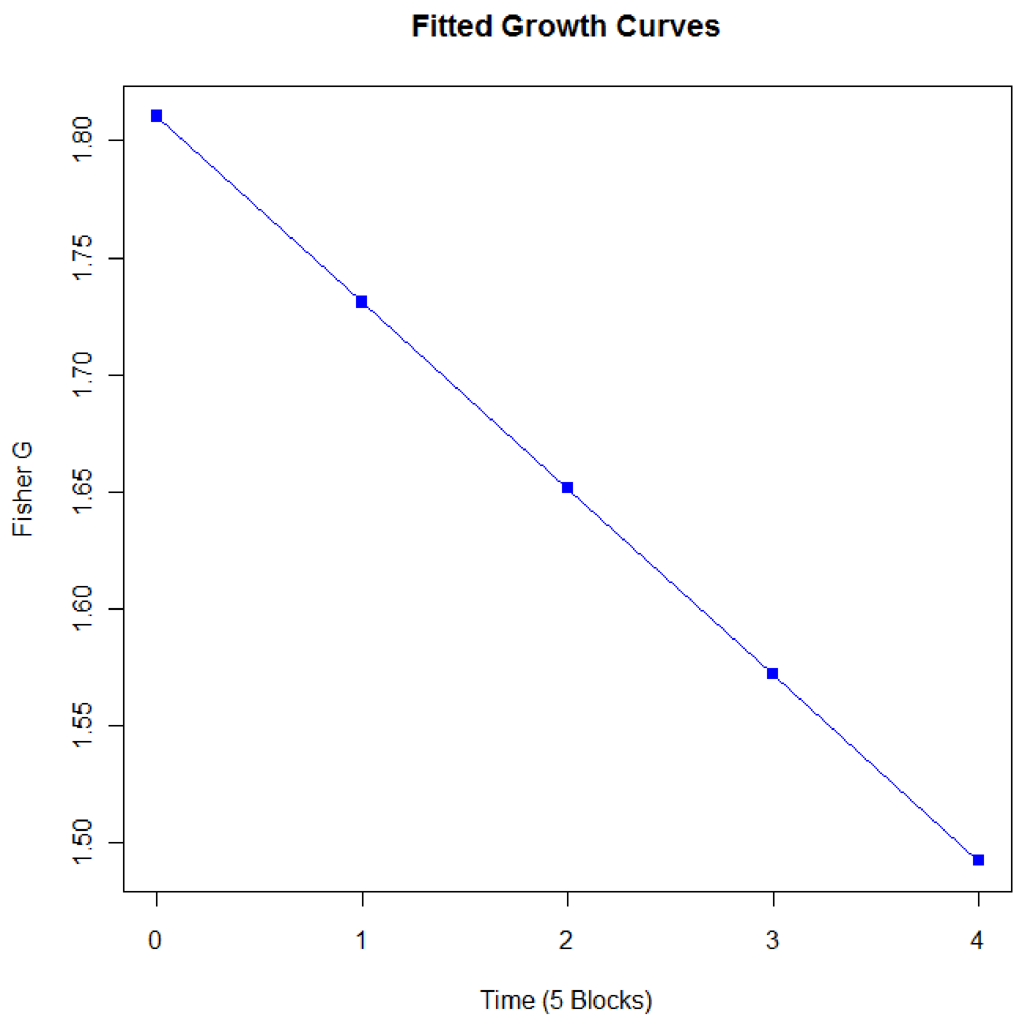


Figure 64. Change Over Time in Fisher R_s (Cognitive Control): Trellis Plot of Slopes and Intercepts (N = All 141 Subjects) (Spring 2010)

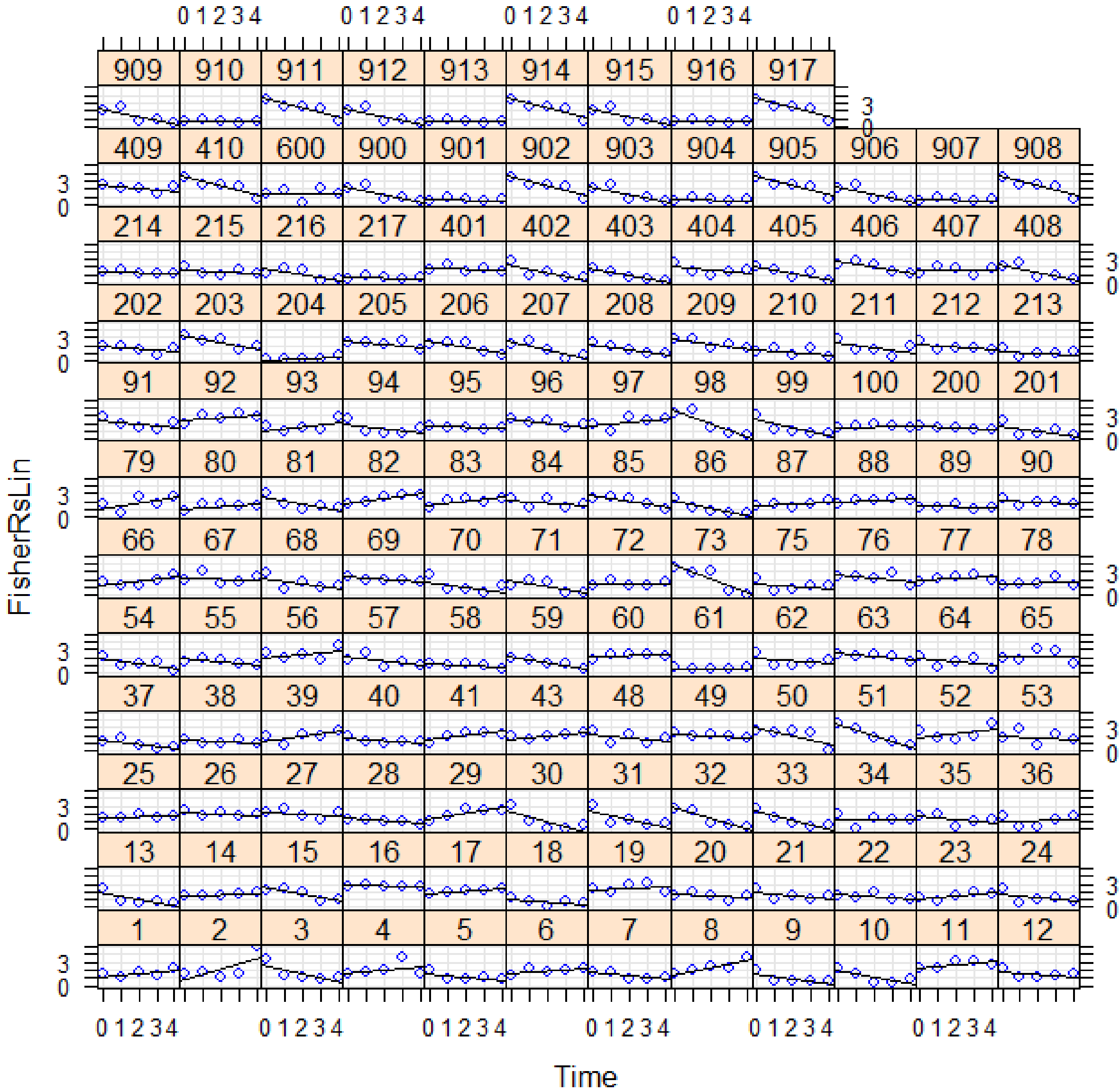


Figure 65. Change Over Time in Fisher R_s (Cognitive Control): Boxplots of Slopes and Intercepts for Feedback Conditions (C1, C2, and C3) (Spring 2010)

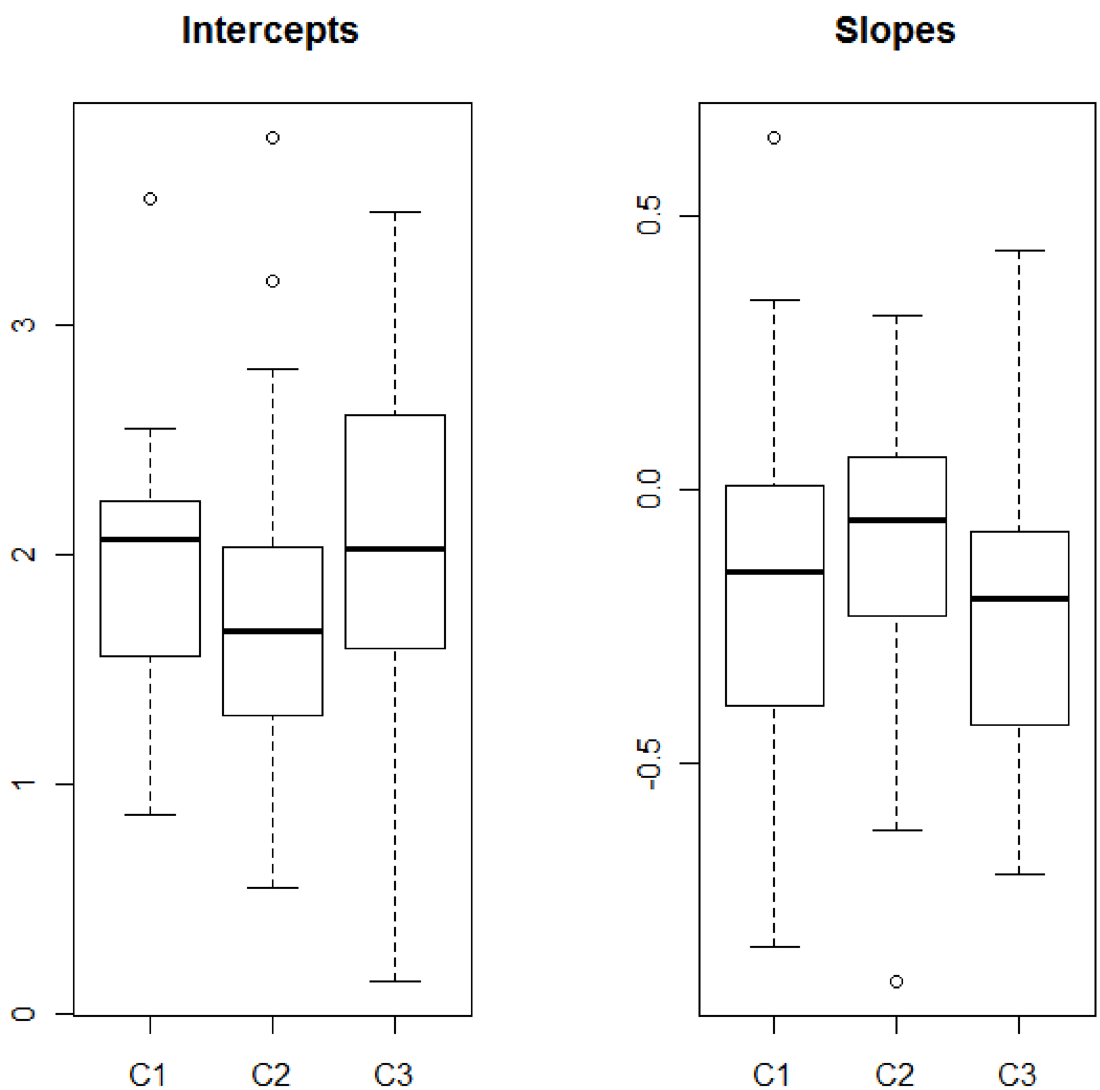


Figure 66. Change Over Time in Fisher R_s (Cognitive Control): Fitted Growth Curves (Spring 2010)

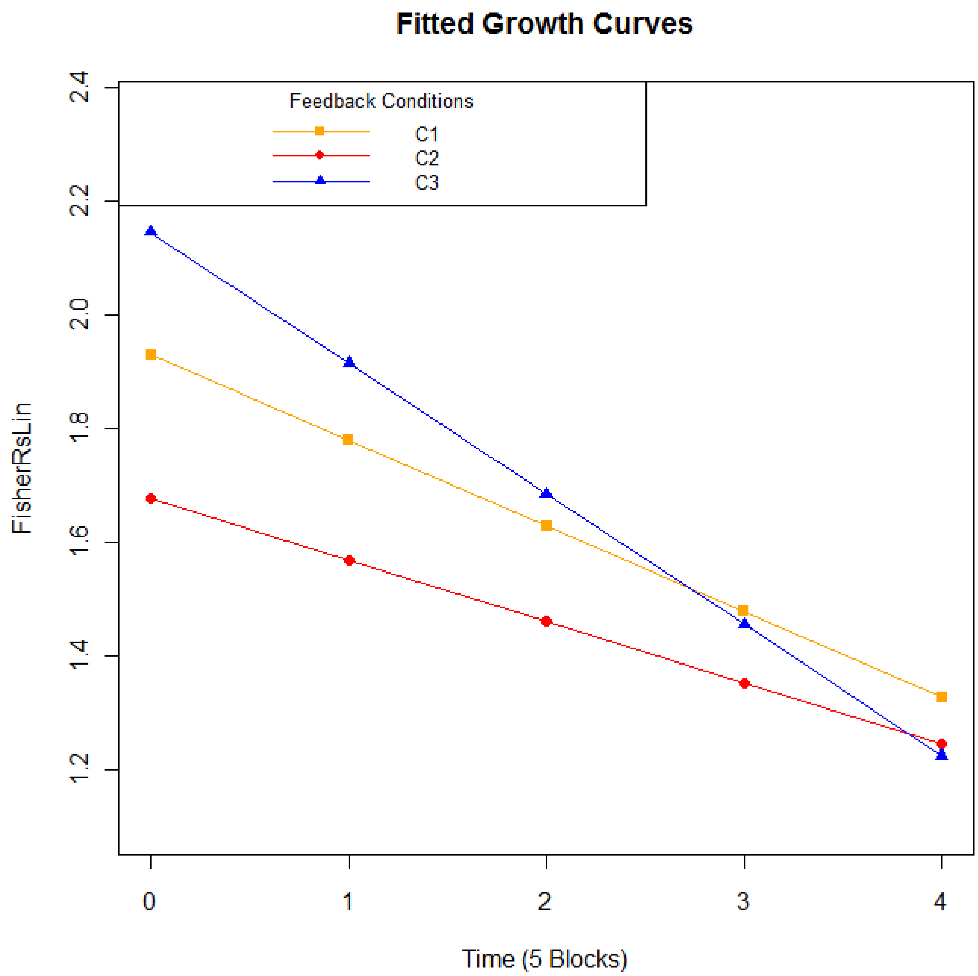


Figure 67. Change Over Time in C_{xj} (Relative Weight for the Disordinal Interaction): Trellis Plot of Slopes and Intercepts (N = All 141 Subjects)

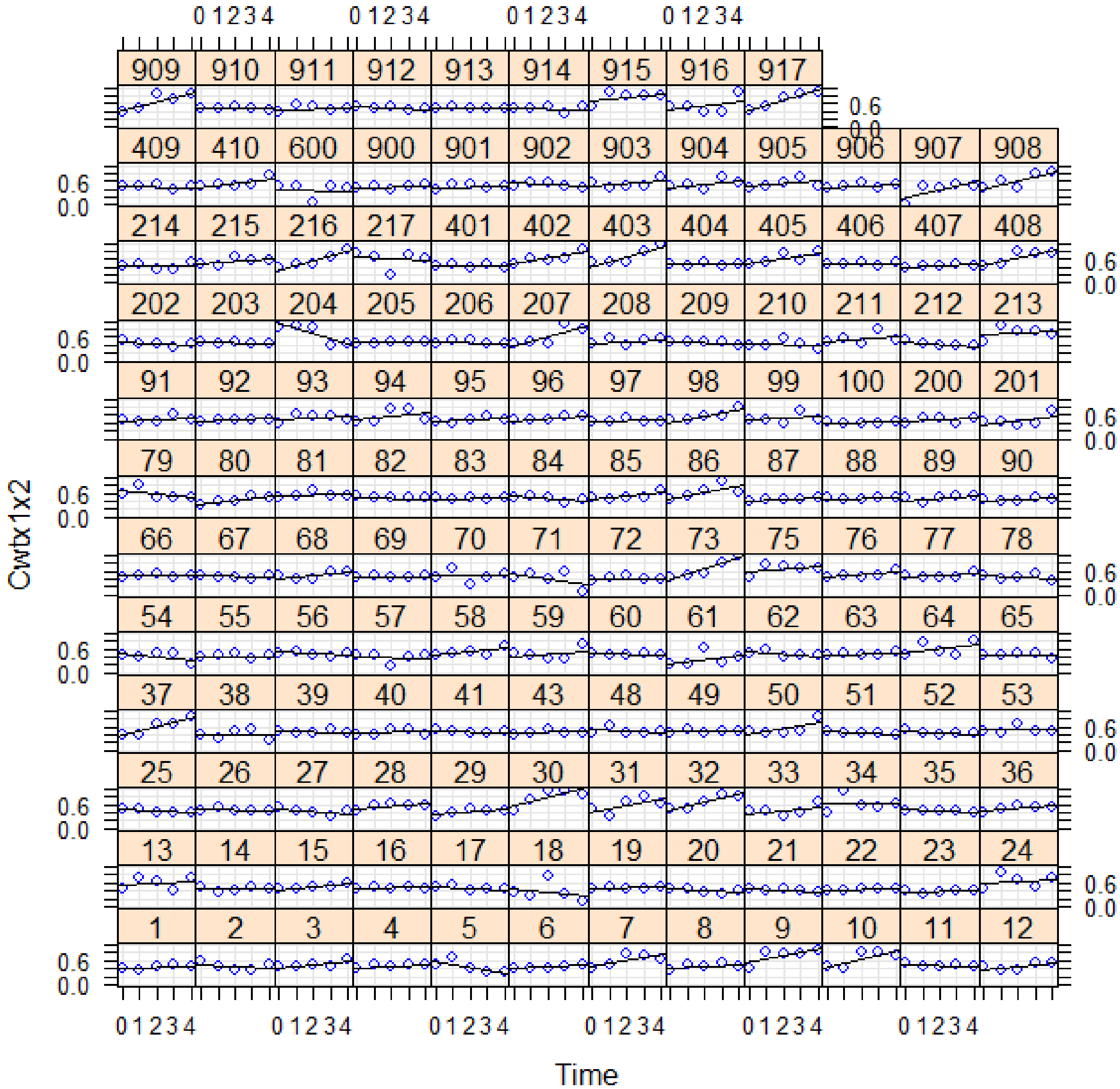


Figure 68. Change Over Time in C_{xj} (Relative Weight for the Disordinal Interaction): Boxplots of Slopes and Intercepts for Feedback Conditions (C1, C2, and C3) (Spring 2010)

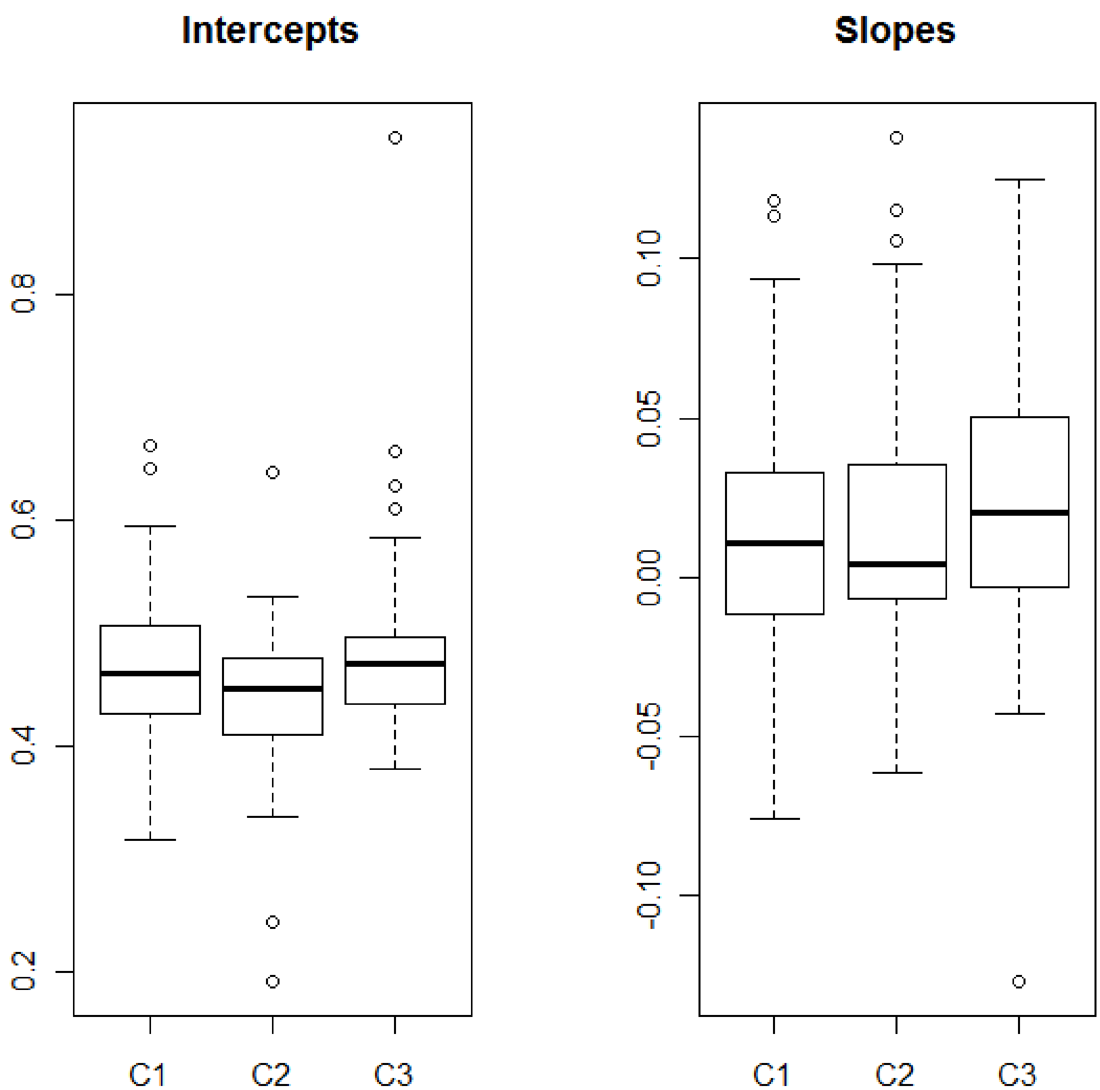


Figure 69. Change Over Time in C_{xj} (Relative Weight for the Disordinal Interaction): Fitted Growth Curves (Spring 2010)

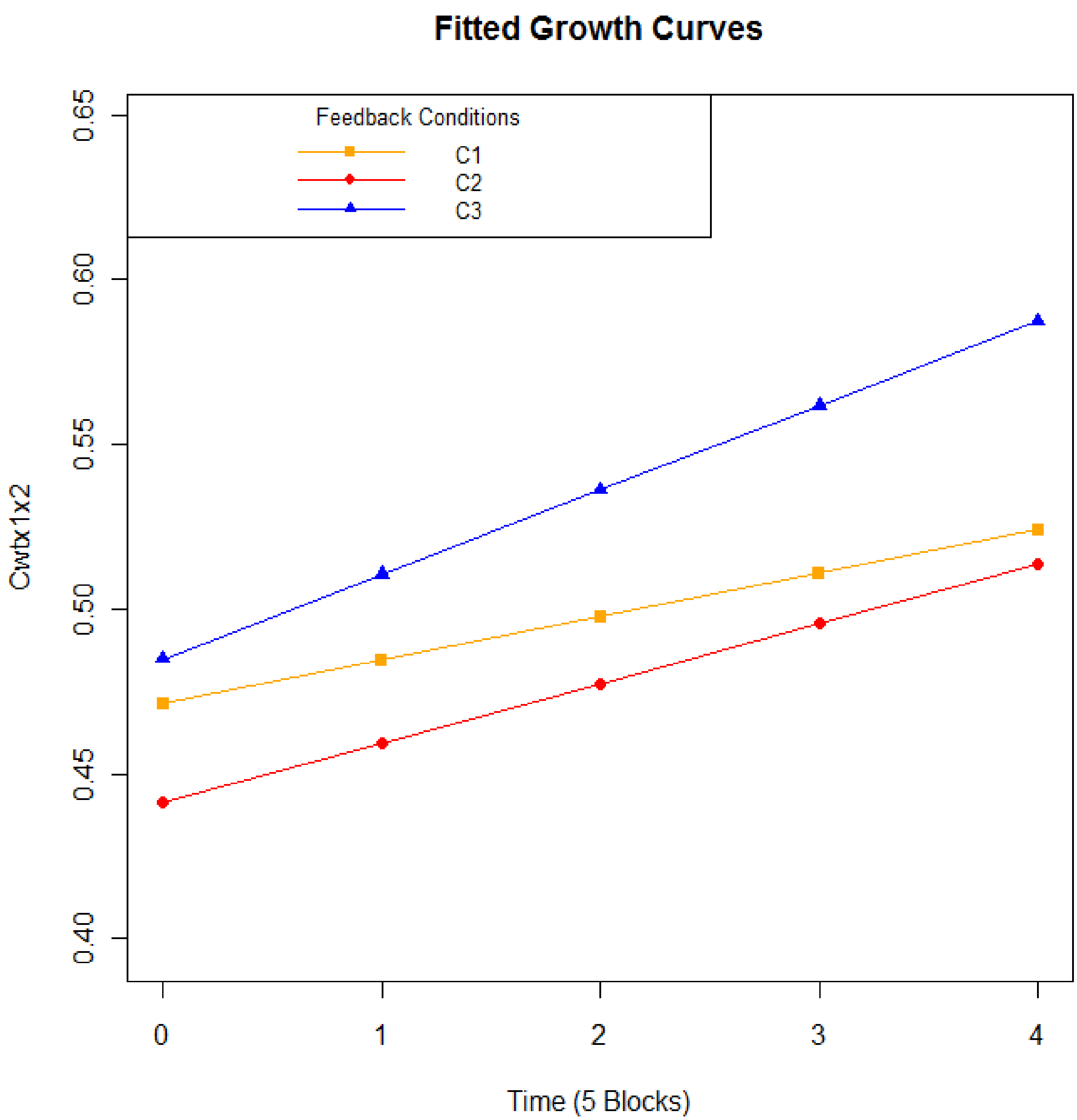


Figure 70. Change Over Time in Mean Absolute Confidence Levels for Each Feedback Condition (Based on Observed Sample-Level Data) (Spring 2010)

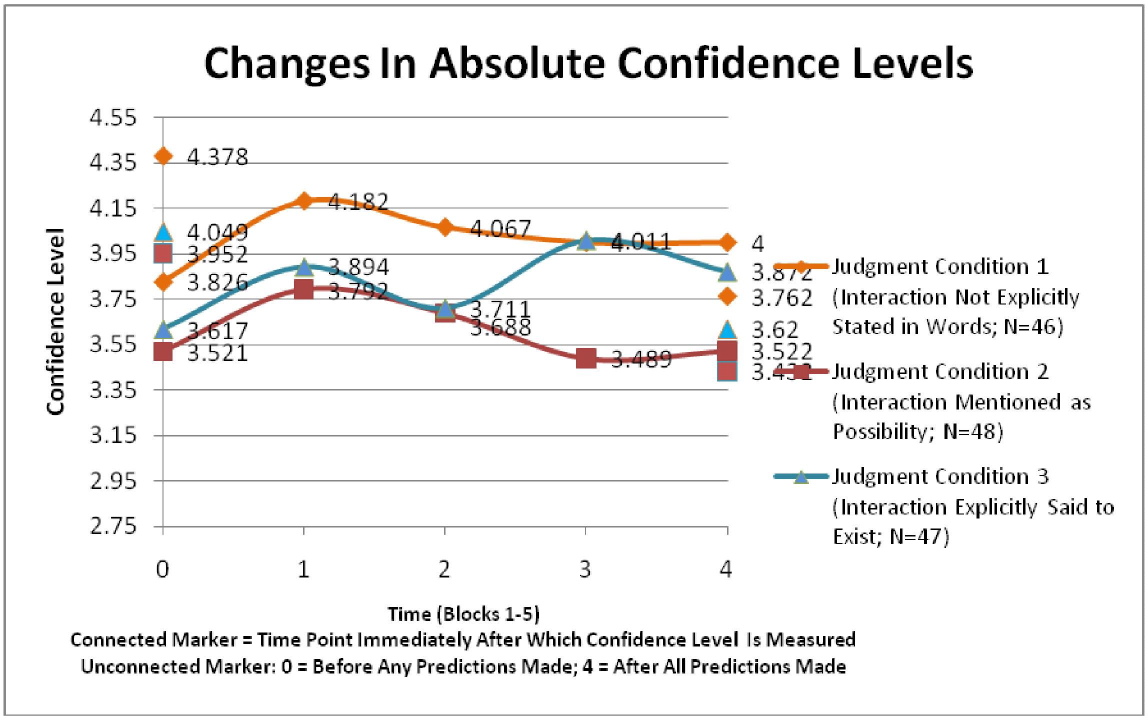


Figure 71. Change Over Time in Mean Relative Confidence Levels for Each Feedback Condition (Based on Observed Sample-Level Data) (Spring 2010)

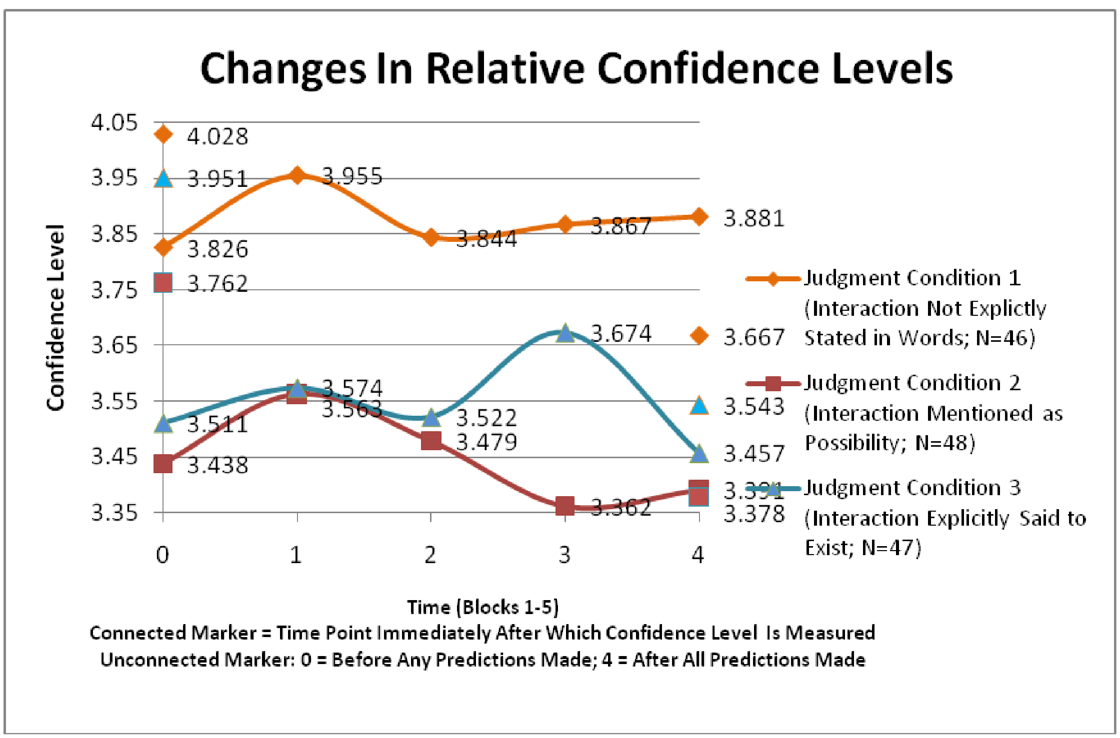


Figure 72. Change Over Time in Absolute Confidence: Trellis Plot of Slopes and Intercepts (N = 140 of 141 Subjects) (Spring 2010)

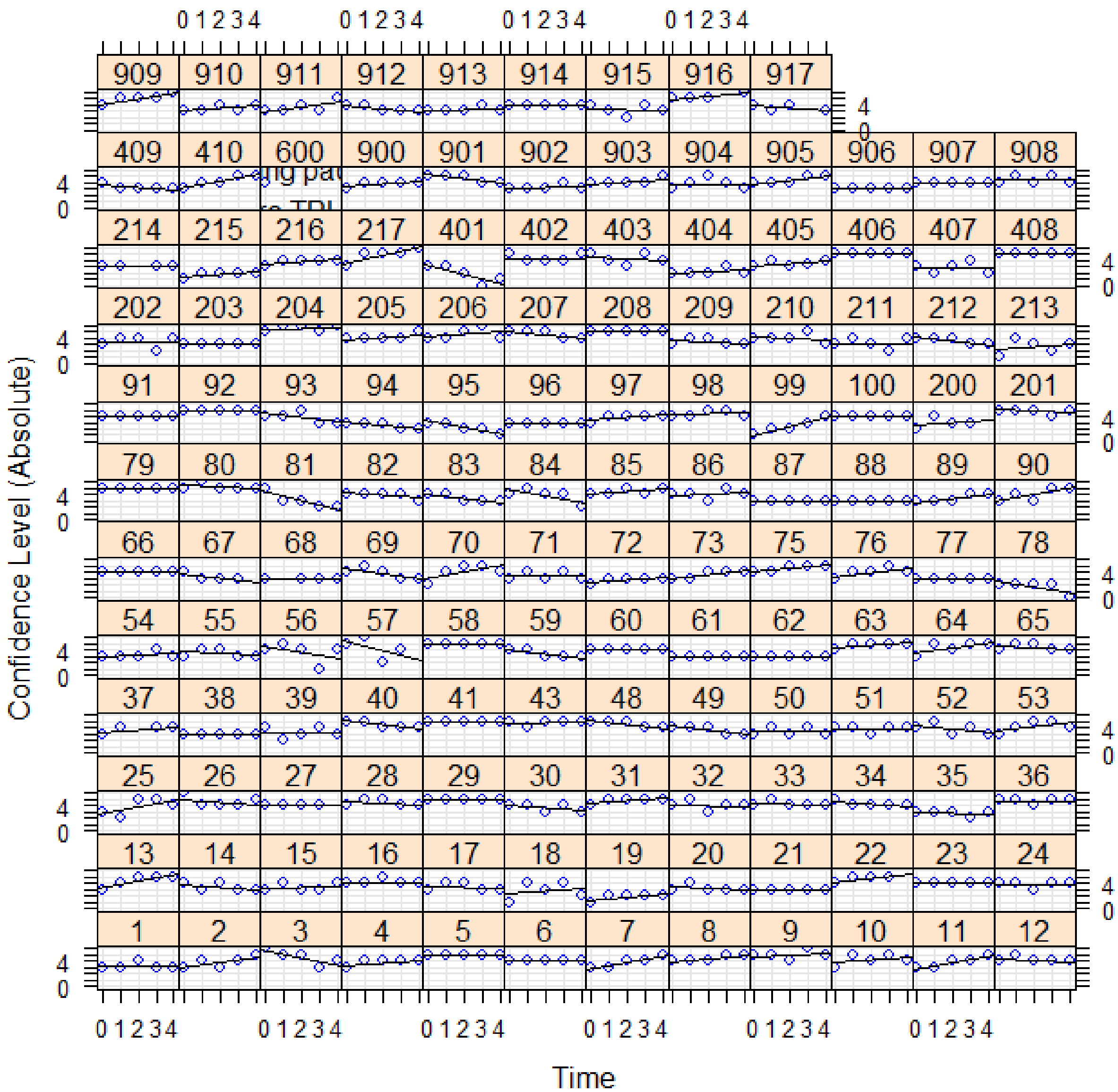


Figure 73. Change Over Time in Absolute Confidence: Boxplots of Slopes and Intercepts for Feedback Conditions (C1, C2, and C3) (Spring 2010)

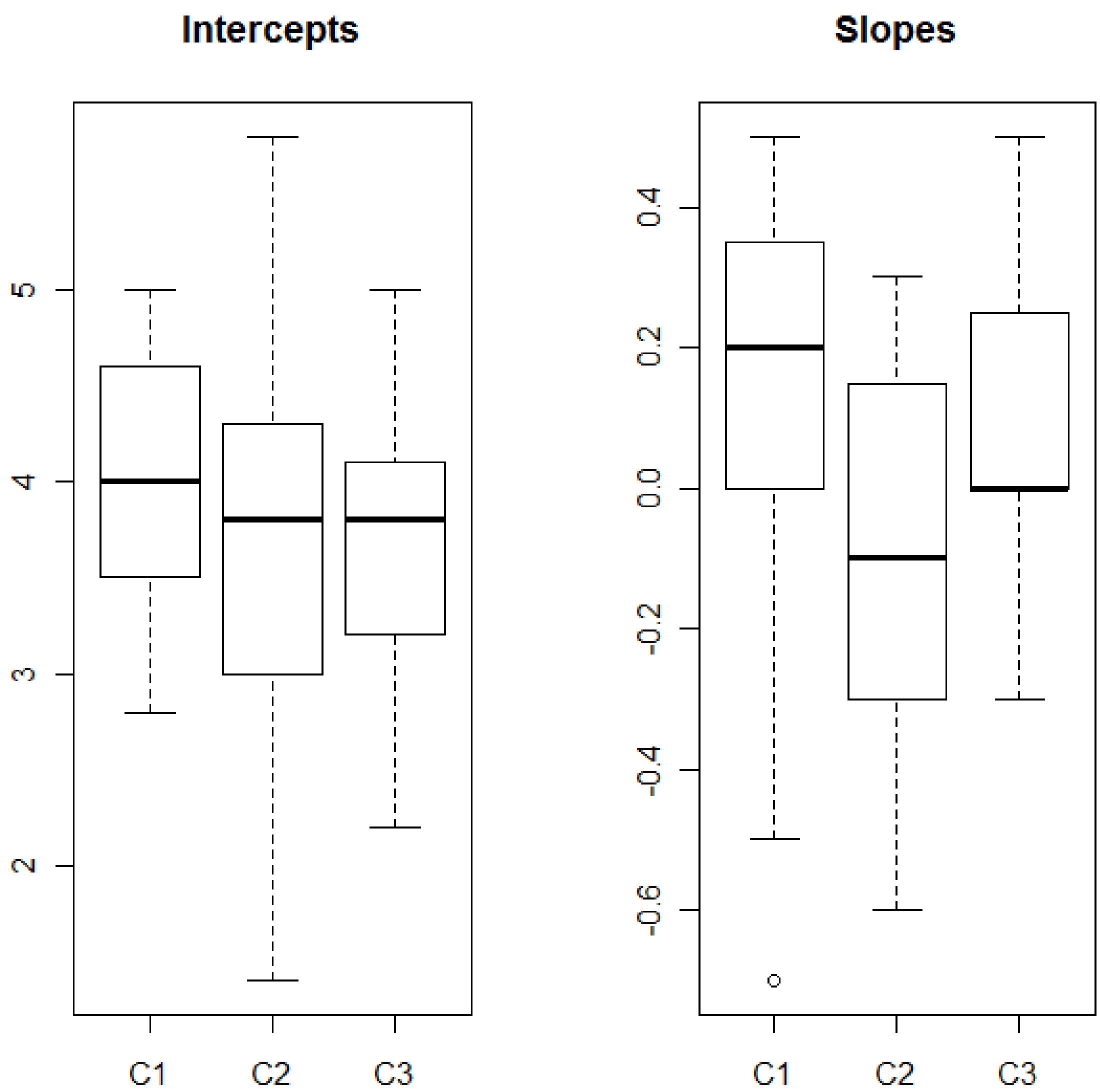


Figure 74. Change Over Time in Absolute Confidence: Fixed Effects Regression Line (No Random Effects; Without Interaction) (N = 140 of 141 Subjects) (Spring 2010)

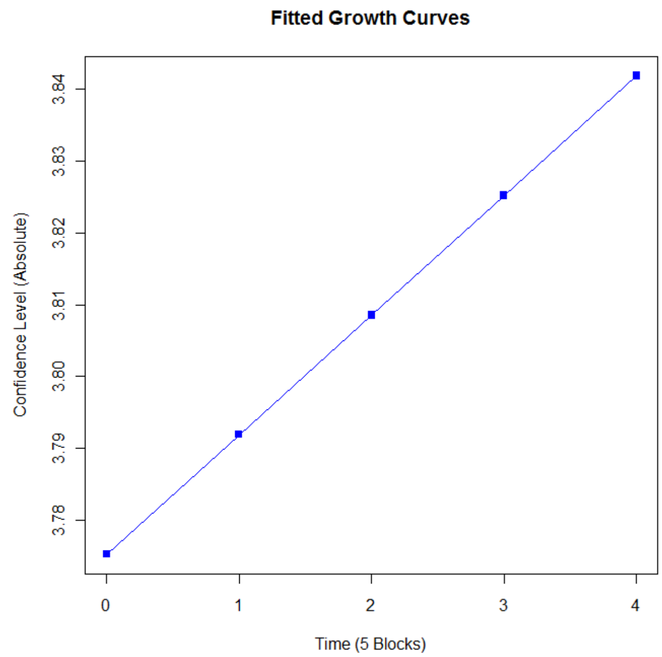


Figure 75. Change Over Time in Absolute Confidence: Fitted Growth Curve for Each Feedback Group (Spring 2010)

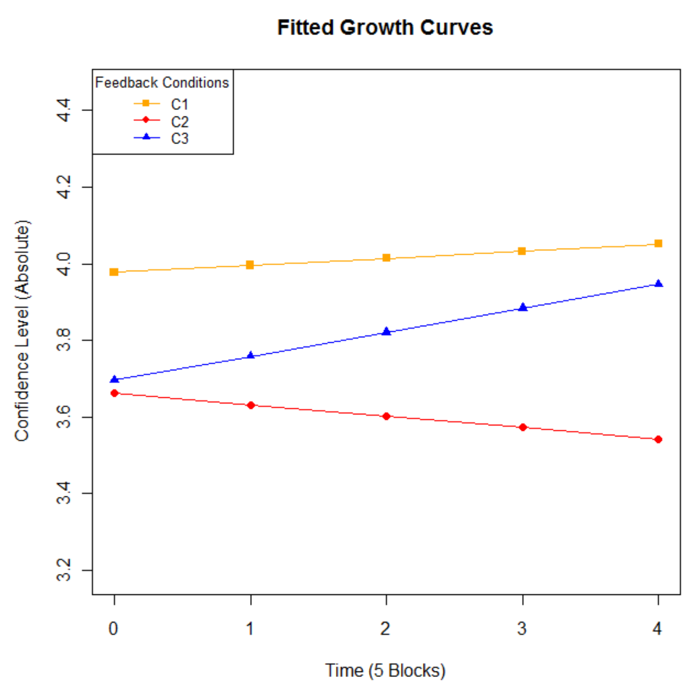


Figure 76. Change Over Time in Relative Confidence: Trellis Plot of Slopes and Intercepts (*N* = 140 of 141 Subjects) (Spring 2010)

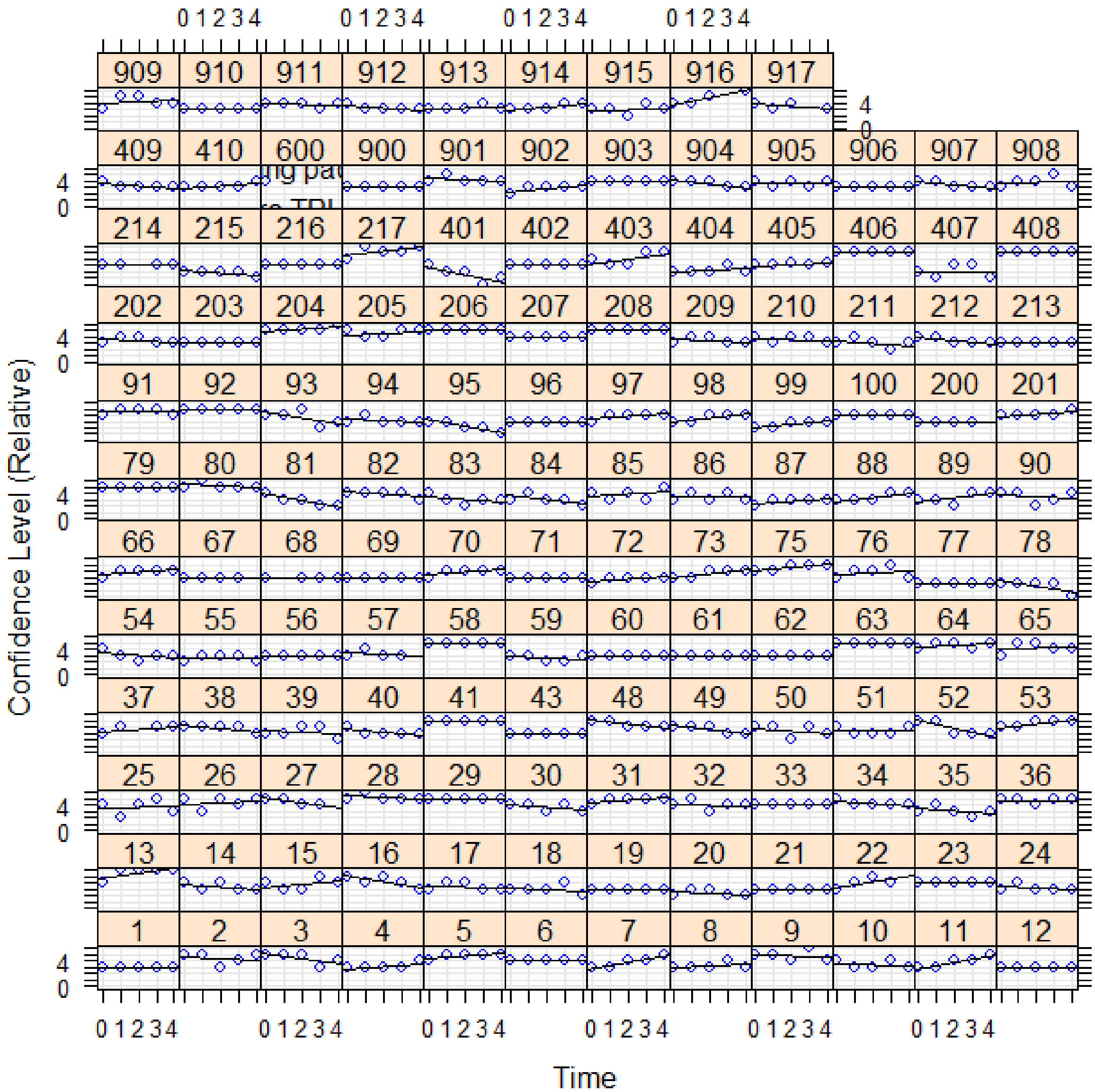


Figure 77. Change Over Time in Relative Confidence: Boxplots of Slopes and Intercepts for Feedback Conditions (C1, C2, and C3) (Spring 2010)

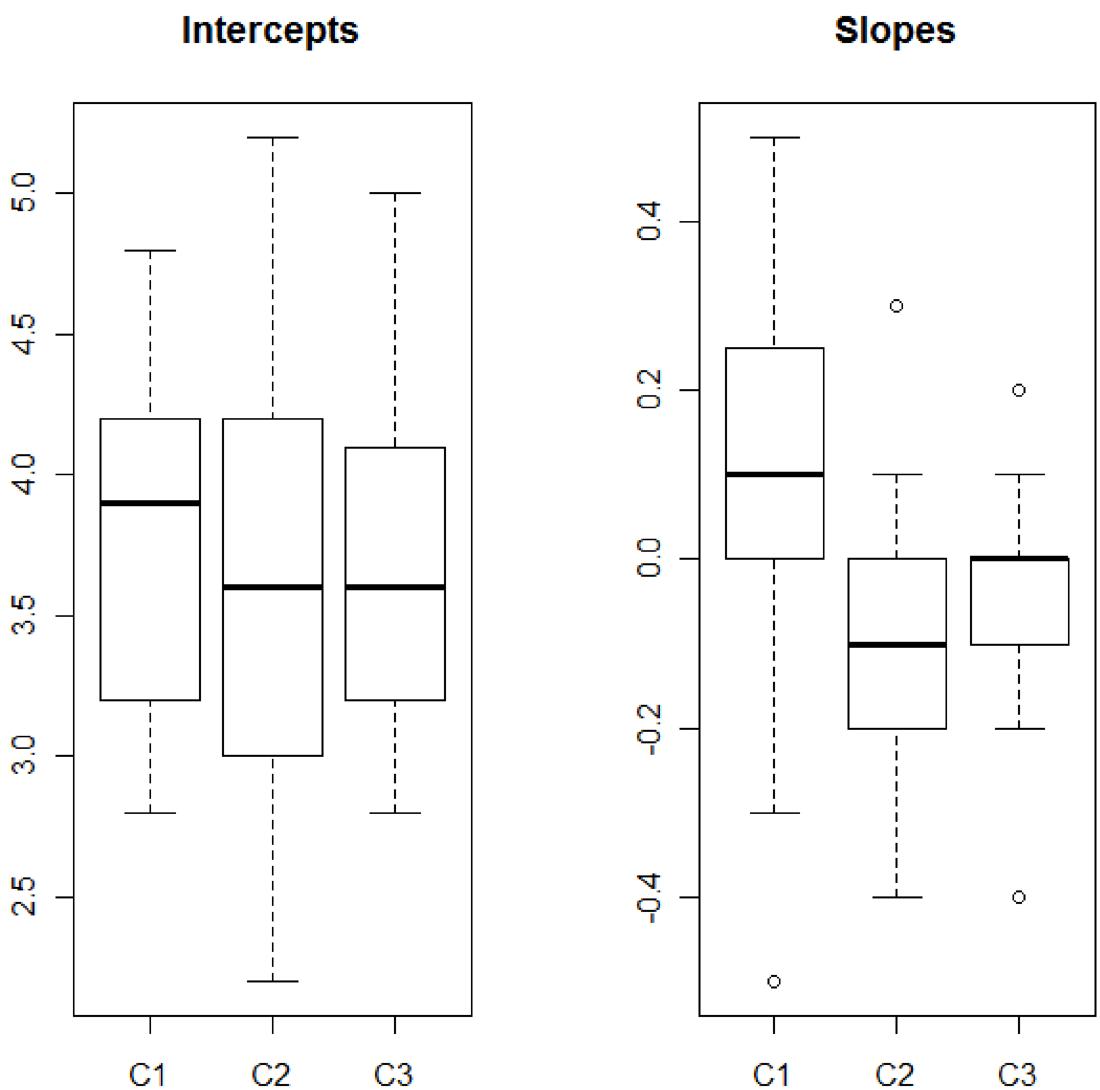


Figure 78. Change Over Time in Relative Confidence: Fixed Effects Regression Line (No Random Effects; Without Interaction) (*N* = 140 of 141 Subjects) (Spring 2010)